Received: 22 June 2023 | Revised: 29 March 2024 | Accepted: 3 April 2024

DOI: 10.1111/sjos.12714

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Asymptotic properties of resampling-based processes for the average treatment effect in observational studies with competing risks

Jasmin Rühl[1] | **Sarah Friedrich**[1,2]

[1]Mathematical Statistics and Artificial Intelligence in Medicine, Augsburg University, Augsburg, Germany

[2]Centre for Advanced Analytics and Predictive Sciences (CAAPS), Augsburg University, Augsburg, Germany

Correspondence
Jasmin Rühl, Mathematical Statistics and Artificial Intelligence in Medicine, Augsburg University, Augsburg, Germany.
Email: jasmin.ruehl@math.uni-augsburg.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: FR 4121/2-1

Abstract
In observational studies with time-to-event outcomes, the g-formula can be used to estimate a treatment effect in the presence of confounding factors. However, the asymptotic distribution of the corresponding stochastic process is complicated and thus not suitable for deriving confidence intervals or time-simultaneous confidence bands for the average treatment effect. A common remedy are resampling-based approximations, with Efron's nonparametric bootstrap being the standard tool in practice. We investigate the large sample properties of three different resampling approaches and prove their asymptotic validity in a setting with time-to-event data subject to competing risks. The usage of these approaches is demonstrated by an analysis of the effect of physical activity on the risk of knee replacement among patients with advanced knee osteoarthritis.

KEYWORDS
average treatment effect, g-formula, resampling, time-to-event data

1 | INTRODUCTION

In observational studies, comparisons between treatment groups are complicated by the potentially unequal distribution of confounding factors. A prominent idea to tackle this issue is the potential outcomes approach, which models the mean outcome in a hypothetical world where all study participants are subject to the same intervention (Hernán & Robins, 2020; Rubin, 1974). This article focuses on the comparison between two treatment groups, where the outcome is a right-censored time-to-event that is possibly subject to competing risks. Our parameter of interest is the average treatment effect (ATE) at time t, defined as the difference between the absolute risks of the event of interest in the exposure groups. Following Ozenne et al. (2020), we model cause-specific hazards by Cox regression models to estimate the absolute risk of the event of interest (Benichou & Gail, 1990; Ozenne et al., 2017).

For thorough statistical inference, confidence intervals and time-simultaneous confidence bands for the ATE provide additional insight. Inference is usually based on Efron's nonparametric bootstrap, since the stochastic processes involved in the estimation are rather complex, though (Efron, 1981; Neumann & Billionnet, 2016; Ryalen et al., 2020; Stensrud et al., 2022). Ozenne et al. (2020). present an alternative approach, which is based on the influence function and the resampling scheme developed by Scheike and Zhang (2008). Another popular resampling technique for time-to-event data is the martingale-based wild bootstrap. This method has successfully been applied in different situations, e.g. in non-causal investigations that cover Cox proportional hazards models (Lin et al., 1993; Lin et al., 1994) or competing risks (Lin, 1997). Extensions that improve the performance for small sample sizes have been discussed by Beyersmann et al. (2013), Dobler and Pauly (2014) as well as Dobler et al. (2017). What is more, the wild bootstrap has been shown to perform superior to the classical bootstrap in several situations, in particular when the data involve dependencies (Nießl et al., 2023; Rühl et al., 2023).

In this paper, we derive a martingale representation of the stochastic process characterizing the asymptotic behaviour of the ATE. Based on this representation, we provide proofs of the asymptotic validity of three resampling approaches: Efron's bootstrap, a resampling method based on the influence function and the wild bootstrap. Thus, the main contribution of this paper is to fill the gap between theory and practice and provide the missing proofs that justify the application of resampling techniques in the situation discussed here.

The remainder of this manuscript is organized as follows: Section 2 introduces the notation, the competing risks setting and the parameter of interest. In Section 3, we investigate the asymptotic behaviour of the estimated ATE and present a martingale-based representation of the corresponding stochastic process, which provides the basis for Section 4, where we establish the asymptotic validity of the three resampling approaches. The corresponding proofs can be found in Section 5. Subsequently, Section 6 summarizes the benefits and drawbacks of the resampling methods. Their application is illustrated by an analysis of publicly accessible data from the Osteoarthritis Initiative (OAI). We close with a discussion in Section 7.

2 | SETTING AND NOTATION

Consider an independent and identically distributed (i.i.d.) data sample of the form $\{(T_i \wedge C_i, D_i, A_i, \boldsymbol{Z_i})\}_{i\in\{1,..,n\}}$. The first vector element denotes the time of an individual's first event (T) or censoring (C), whichever occurs earlier. In a setting with K competing causes of failure, the indicator D may assume values in $\{1,\dots,K\}$ according to the type of event that is observed, whereas for censored observations, $D=0$. The data moreover include the treatment indicator $A\in\{0,1\}$ as well as the covariate vector $\boldsymbol{Z}\in\mathbb{R}^p$. We suppose that there are no ties and that $T \perp\!\!\!\perp C$ conditional on $(A,\boldsymbol{Z})$. Besides, the covariate values in $\boldsymbol{Z}$ should be bounded.

Let without loss of generality $D=1$ refer to the event of interest and define the potential cumulative incidence function $F_1^a(t)=P(T^a\le t,\ D^a=1)$ under treatment a. This function quantifies the probability of experiencing the event of interest until time t in a hypothetical world where all individuals received treatment a. Given independent censoring, that probability can be represented by observable data alone if the cause-specific hazards are predictable (i.e., nonrandom or dependent only on past events and covariates known by the researcher):

$$F_1^a(t) = \int_0^t \mathbb{1}\left\{T^a \wedge C^a \geq s\right\} \cdot \exp\left(-\sum_{k=1}^{K} \int_0^s P\left(T^a \wedge C^a \in [u, u+\mathrm{d}u),\ D^a = k \middle| T^a \wedge C^a \geq u\right) \mathrm{d}u \right) \cdot P\left(T^a \wedge C^a \in [s, s+\mathrm{d}s),\ D^a = 1 \middle| T^a \wedge C^a \geq s\right) \mathrm{d}s.$$

The expression $\mathbb{1}\{\cdot\}$ above is used to denote the indicator function of the event in the argument. To compare the effectiveness of two different treatment strategies $a = 1$ and $a = 0$, we consider F_1^a as estimand and characterize the ATE by the relation $\mathrm{ATE}(t) = \mathbb{E}\left(F_1^1(t) - F_1^0(t)\right)$, with time t ranging between 0 and τ, the terminal time of the study. Note that ATE merely describes the total effect on the event of interest, that is, one cannot distinguish between the direct influence of treatment on the event of interest and the impact that is due to advancing/preventing competing events ts (Stensrud et al., 2022). If the identifiability conditions of exchangeability, positivity and consistency apply ply (cf. Hernán & Robins, 2020, Sec. I.3), the g-formula suggests

$$\widehat{\mathrm{ATE}}(t) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{F}_1\left(t \middle| A = 1, \boldsymbol{Z_i}\right) - \hat{F}_1\left(t \middle| A = 0, \boldsymbol{Z_i}\right) \right),$$

as an estimator of the ATE (Ozenne et al., 2020). One possible way to obtain $\hat{F}_1\left(t \middle| A, \boldsymbol{Z}\right)$, i.e. an estimator for the cumulative incidence function of the cause of interest given treatment A and covariates $\boldsymbol{Z}$, involves fitting cause-specific Cox models with covariates A and $\boldsymbol{Z^{(k)}} \in \mathbb{R}^{p_k}$ for each event type $k \in \{1, \ldots, K\}$. This yields estimated cumulative hazards of the following form:

$$\hat{\Lambda}_k\left(t \middle| a, \boldsymbol{z^{(k)}}\right) = \hat{\Lambda}_{0k}(t)\ \exp\left(\hat{\beta}_{kA} a + \hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{z^{(k)}}\right),$$
$$\text{with } \hat{\Lambda}_{0k}(t) = \int_0^t \frac{\mathrm{d}N_k(s)}{\sum_{i=1}^n Y_i(s)\ \exp\left(\hat{\beta}_{kA} A_i + \hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{Z_i^{(k)}}\right)}.$$

(We use Z instead of $Z^{(k)}$ hereafter, as the cause specificity of the covariates follows from the context.) The vector $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} = (\hat{\beta}_{kA}, \hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T)^T \in \mathbb{R}^{p_k+1}$ results from the Cox regression and combines the estimated coefficients for treatment and covariates. Apart from that, $\hat{\Lambda}_{0k}(t)$ is the Breslow estimator of the cumulative baseline hazard (Breslow, 1972), which depends on the counting process $N_k(t) = \sum_{i=1}^n N_{ki}(t)$ for observed events of type k, with $N_{ki}(t) = \mathbb{1}\{T_i \wedge C_i \leq t,\ D_i = k\}$, and the at-risk indicator $Y_i(t) = \mathbb{1}\{T_i \wedge C_i \geq t\},\ i \in \{1, \ldots, n\}$. The estimator of the cumulative incidence finally results by plugging the estimated cumulative hazard $\hat{\Lambda}_k\left(t \middle| a, \boldsymbol{z}\right)$, into the formula

$$\hat{F}_1\left(t \middle| a, \boldsymbol{z}\right) = \int_0^t \hat{S}\left(t \middle| a, \boldsymbol{z}\right)\ \mathrm{d}\hat{\Lambda}_1\left(s \middle| a, \boldsymbol{z}\right).$$

Here, $\hat{S}\left(t\middle|a,\boldsymbol{z}\right) = \exp\left(-\sum_{k=1}^{K}\hat{\Lambda}_k\left(s\middle|a,\boldsymbol{z}\right)\right)$ approximates the survival probability $P\left(T > t\middle|a,\boldsymbol{z}\right)$ for a given treatment a and covariate vector $\boldsymbol{z}$.

3 | ASYMPTOTIC DISTRIBUTION OF THE ESTIMATED ATE

In order to investigate the asymptotic behaviour of $\widehat{\text{ATE}}$, we study the process $U_n(t) = \sqrt{n}\left(\widehat{\text{ATE}}(t) - \text{ATE}(t)\right)$ and its properties as $n \to \infty$. Arguments similar to those used by Cheng et al. (1998) show that the limiting distribution of U_n may be represented in terms of martingales. This is an important finding that facilitates further inferences on the large-sample properties of the process. Before we introduce the martingale representation of the process, it is necessary to define several functions and variables, however.

Consider the subsequent quantities, which are useful to express the score function and the Fisher information for the Cox model (cf. Andersen et al., 1993, Sec. VII.2, in particular Ex. VII.2.3);

$$S^{(r)}(\boldsymbol{\beta_k}, t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\ \exp\left(\beta_{kA}A_i + \boldsymbol{\beta}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right)\left(\left(A_i, \boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T\right)^{\otimes r}, \quad r \in \{0,1,2\},$$

(with $\boldsymbol{v}^{\otimes 0} = 1$, $\boldsymbol{v}^{\otimes 1} = \boldsymbol{v}$, $\boldsymbol{v}^{\otimes 2} = \boldsymbol{v}\boldsymbol{v}^T$ for a column vector $\boldsymbol{v}$), and the corresponding expectations $s^{(r)}(\boldsymbol{\beta_k}, t) = \mathbb{E}\left(S^{(r)}(\boldsymbol{\beta_k}, t)\right)$, $r \in \{0,1,2\}$, as well as

$$\boldsymbol{E}(\boldsymbol{\beta_k}, t) = \frac{\boldsymbol{S}^{(\boldsymbol{1})}(\boldsymbol{\beta_k}, t)}{S^{(0)}(\boldsymbol{\beta_k}, t)},$$

with $\boldsymbol{e}(\boldsymbol{\beta_k}, t) = \boldsymbol{s}^{(\boldsymbol{1})}(\boldsymbol{\beta_k}, t) \Big/ s^{(0)}(\boldsymbol{\beta_k}, t)$ for $k \in \{1, \dots, K\}$. Let the positive definite matrix $\boldsymbol{\Sigma_k}$ be the inverse covariance matrix of the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$, which is given by

$$\boldsymbol{\Sigma_k} = \int_0^{\tau}\left(\frac{\boldsymbol{s}^{(\boldsymbol{2})}(\boldsymbol{\beta_{0k}}, u)}{s^{(0)}(\boldsymbol{\beta_{0k}}, u)} - \left(\boldsymbol{e}(\boldsymbol{\beta_{0k}}, u)\right)^{\otimes 2}\right)s^{(0)}(\boldsymbol{\beta_{0k}}, u)\ \mathrm{d}\Lambda_{0k}(u),$$

supposing that the Cox model applies with true vector of regression coefficients $\boldsymbol{\beta_{0k}}$ for cause k (cf. Andersen et al., 1993, Thm. VII.2.2). The following functions have further been introduced by Cheng et al. (1998) to characterize the asymptotic distribution of the cumulative incidence:

$$\boldsymbol{h_k}\left(t\middle|a,\boldsymbol{z}\right) = \int_0^t \left(\left(a, \boldsymbol{z}^T\right)^T - \boldsymbol{e}(\boldsymbol{\beta_{0k}}, u)\right)\ \mathrm{d}\Lambda_k\left(u\middle|a,\boldsymbol{z}\right),$$

$$\boldsymbol{\varphi_1}\left(t\middle|a,\boldsymbol{z}\right) = \int_0^t S\left(u-\middle|a,\boldsymbol{z}\right)\ \mathrm{d}\boldsymbol{h_1}\left(u\middle|a,\boldsymbol{z}\right),$$

$$\boldsymbol{\psi_{1k}}\left(t\middle|a,\boldsymbol{z}\right) = \int_0^t \left(F_1\left(t\middle|a,\boldsymbol{z}\right) - F_1\left(u\middle|a,\boldsymbol{z}\right)\right)\ \mathrm{d}\boldsymbol{h_k}\left(u\middle|a,\boldsymbol{z}\right).$$

Eventually, we define

$$H_{k1i}(u,t) = \frac{\tilde{H}_{k1}(u,t)}{\sqrt{n}\ S^{(0)}(\boldsymbol{\beta_{0k}}, u)} \quad \text{and} \quad H_{k2i}(u,t) = \frac{1}{\sqrt{n}}\ \left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\left(A_i, \boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}}, u)\right),$$

$k \in \{1,\ldots,K\}$, $i \in \{1,\ldots,n\}$, $u \leq t$, with

$$\tilde{H}_{11}(u,t) = \frac{1}{n}\sum_{i=1}^{n}\Bigg(\Big(S\Big(u-\Big|A=1,\boldsymbol{Z_i}\Big) - F_1\Big(t\Big|A=1,\boldsymbol{Z_i}\Big) + F_1\Big(u\Big|A=1,\boldsymbol{Z_i}\Big)\Big)\exp\left(\beta_{01A}+\boldsymbol{\beta}_{\boldsymbol{01Z}}^T\boldsymbol{Z_i}\right)$$
$$-\Big(S\Big(u-\Big|A=0,\boldsymbol{Z_i}\Big) - F_1\Big(t\Big|A=0,\boldsymbol{Z_i}\Big) + F_1\Big(u\Big|A=0,\boldsymbol{Z_i}\Big)\Big)\exp\left(\boldsymbol{\beta}_{\boldsymbol{01Z}}^T\boldsymbol{Z_i}\right)\Bigg),$$
$$\tilde{H}_{k1}(u,t) = \frac{1}{n}\sum_{i=1}^{n}\Bigg(\Big(F_1\Big(t\Big|A=0,\boldsymbol{Z_i}\Big) - F_1\Big(u\Big|A=0,\boldsymbol{Z_i}\Big)\Big)\exp\left(\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z_i}\right)$$
$$-\Big(F_1\Big(t\Big|A=1,\boldsymbol{Z_i}\Big) - F_1\Big(u\Big|A=1,\boldsymbol{Z_i}\Big)\Big)\exp\left(\beta_{0kA}+\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z_i}\right)\Bigg), \quad k \in \{2,\ldots,K\},$$

and

$$\tilde{\boldsymbol{H}}_{\boldsymbol{12}}(t) = \frac{1}{n}\sum_{i=1}^{n}\Bigg(\Big(\boldsymbol{\varphi_1}\Big(t\Big|A=1,\boldsymbol{Z_i}\Big) - \boldsymbol{\psi_{11}}\Big(t\Big|A=1,\boldsymbol{Z_i}\Big)\Big) - \Big(\boldsymbol{\varphi_1}\Big(t\Big|A=0,\boldsymbol{Z_i}\Big) - \boldsymbol{\psi_{11}}\Big(t\Big|A=0,\boldsymbol{Z_i}\Big)\Big)\Bigg),$$
$$\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big(\boldsymbol{\psi_{1k}}\Big(t\Big|A=0,\boldsymbol{Z_i}\Big) - \boldsymbol{\psi_{1k}}\Big(t\Big|A=1,\boldsymbol{Z_i}\Big)\Big), \quad k \in \{2,\ldots,K\}.$$

Based on these functions, one finds an adjuvant approximation of U_n:

Lemma 1. *For the process*

$$\tilde{U}_n(t) = \sum_{k=1}^{K}\sum_{i=1}^{n}\left(\int_0^t H_{k1i}(s,t)\,\mathrm{d}M_{ki}(s) + \int_0^\tau H_{k2i}(s,t)\,\mathrm{d}M_{ki}(s)\right),$$

with $M_{ki}(t) = N_{ki}(t) - \int_0^t Y_i(s)\,\mathrm{d}\Lambda_k\Big(s\Big|A_i,\boldsymbol{Z_i}\Big)$, $k \in \{1,\ldots,K\}$, $i \in \{1,\ldots,n\}$, *it holds that*

$$U_n(t) = \tilde{U}_n(t) + o_p(1).$$

Note that M_{ki} is a martingale relative to the history $(\mathscr{F}_t)_{t\geq 0}$ that is generated by the data observed until a given time, that is, $\mathbb{E}\Big(\mathrm{d}M_{ki}(t)\Big|\mathscr{F}_{t-}\Big) = 0$. The proofs of this and all following propositions are deferred to Section 5 for better readability.

The subsequent theorem characterizes the limiting distribution of U_n for fixed covariate vectors $\boldsymbol{Z_i}$, $i \in \{1,\ldots,n\}$:

Theorem 1. *The process U_n converges weakly to a zero-mean Gaussian process with covariance function* $\xi(t_1,t_2) = \sum_{k=1}^{K}\xi^{(k)}(t_1,t_2)$,

$$\xi^{(k)}(t_1,t_2) = \int_0^{t_1\wedge t_2}\tilde{H}_{k1}(u,t_1)\,\tilde{H}_{k1}(u,t_2)\,\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)} + \left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_1)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\,\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_2),$$

on the Skorokhod space $\mathcal{D}[0,\tau]$.

4 | RESAMPLING-BASED APPROXIMATIONS

The asymptotic distribution of $U_n(t)$ is too complex to derive in practice, which is why resampling approaches are often used as a remedy to draw inferences on $\mathrm{ATE}(t)$. In the following, we show the validity of three different methods.

4.1 | Efron's bootstrap

Usually, confidence intervals and bands for the ATE are constructed using the classical nonparametric bootstrap (Efron, 1981). The main idea is to draw n times with replacement from the data at hand and compute the desired statistical functional in the resulting bootstrap sample. This step is repeated multiple times, yielding a set of bootstrap estimators that provides information on the distribution of the underlying functional. Although this approach generally provides asymptotically valid outcomes, there are certain situations where it breaks down (Friedrich et al., 2017; Singh, 1981). To the best of our knowledge, a proof of the validity in the specific setting considered here is still pending.

Theorem 2.

$$U_n^*(t) = \sqrt{n}\left(\widehat{\text{ATE}}^*(t) - \widehat{\text{ATE}}(t)\right),$$

with $\widehat{\text{ATE}}^*(t)$ *being the estimated ATE in the bootstrap sample, converges to the same limiting process as* $U_n(t)$ *for almost all data samples* $\{T_i \wedge C_i, D_i, A_i, \boldsymbol{Z_i}\}_{i\in\{1,..,n\}}$ *if* $\inf_{u\in[0,\tau]} Y(u) \xrightarrow{P} \infty$.

4.2 | Influence function

Ozenne et al. (2020) presented a second resampling technique based on the influence function of the ATE. The idea proceeds from the functional delta method, which shows that

$$U_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\text{IF}(t;\ T_i \wedge C_i, D_i, A_i, \boldsymbol{Z_i}) + o_P(1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \int (IF(t;\ s,d,a,z))^2\, \mathrm{d}P(s,d,a,z)\right).$$

To look up the definition of the influence function IF, refer to Ozenne et al. (2020, 2017). The authors propose the resampling method described by Scheike and Zhang (2008) in order to approximate the distribution of the process while taking the dependence of the increments of U_n into account. This method is valid because of the asymptotic properties of U_n (see Theorem 1). More specifically, one can imitate the limiting distribution of U_n by applying independent standard normal variables $G_1^{\text{IF}}, \ldots, G_n^{\text{IF}}$ and the plug-in estimator $\widehat{\text{IF}}$ as follows:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\text{IF}}(t;\ T_i \wedge C_i, D_i, A_i, \boldsymbol{Z_i}) \cdot G_i^{\text{IF}}.$$

For further details, see also van der Vaart (1998, Chap. 20).

4.3 | Wild bootstrap

With regard to the martingale representation of U_n from Lemma 1, a third approach results in accordance with the resampling scheme proposed by Lin et al. (1993). In short, one tries to emulate the distribution of the martingale increments $\mathrm{d}M_{ki}, k \in \{1,\ldots,K\}, i \in \{1,\ldots,n\}$, by generating random variates with asymptotically equal moments. The subsequent theorem sets out the conditions these variates need to fulfil in more detail. (Note the parallels to Theorem 1 in Dobler et al., 2017.)

Theorem 3. *Let G_i^{WB}, $i \in \{1, \dots, n\}$, be random variables that satisfy the following conditions*:

(i) $\sqrt{n} \max_{1 \le i \le n} \mathbb{E}\left(G_i^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right) \xrightarrow{P} 0;$

(ii) $\max_{1 \le i \le n} \mathrm{Var}\left(G_i^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right) \xrightarrow{P} 1;$

(iii) $1 \Big/ \sqrt{n} \max_{1 \le i \le n} \mathbb{E}\left(\left(G_i^{\mathrm{WB}}\right)^4 \middle| \mathscr{F}_\tau\right) \xrightarrow{P} 0;$

(iv) $\mathcal{L}\left(G_1^{\mathrm{WB}}, \dots, G_n^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right) = \otimes_{i=1}^n \mathcal{L}\left(G_i^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right),$

where $\mathcal{L}\left(\cdot \middle| \mathscr{F}_\tau\right)$ denotes the conditional distribution given $\mathscr{F}_\tau$, and $\otimes$ is the product measure;

(v) $\sum_{i=1}^n \mathbb{E}\left(\frac{\left(G_i^{\mathrm{WB}} - \mathbb{E}\left(G_i^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right)\right)^2}{\sum_{j=1}^n \left(\mathrm{Var}\left(G_j^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right)\right)^2} \cdot \mathbb{1}\left\{ \frac{\left(G_i^{\mathrm{WB}} - \mathbb{E}\left(G_i^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right)\right)^2}{\sum_{j=1}^n \left(\mathrm{Var}\left(G_j^{\mathrm{WB}} \middle| \mathscr{F}_\tau\right)\right)^2} > \epsilon \right\} \middle| \mathscr{F}_\tau \right) \xrightarrow{P} 0 \ \forall \epsilon > 0.$

Then the plug-in estimate of U_n,

$$\hat{U}_n(t) = \sum_{k=1}^{K} \sum_{i=1}^{n} \left(\hat{H}_{k1i}(T_i \wedge C_i, t)\, N_{ki}(t) G_i^{\mathrm{WB}} + \hat{H}_{k2i}(T_i \wedge C_i, t)\, N_{ki}(\tau) G_i^{\mathrm{WB}} \right),$$

converges weakly to the same process as U_n on $\mathcal{D}[0, \tau]$, conditional on the data.

The functions $\hat{H}_{k1i}$ and $\hat{H}_{k2i}$ are calculated by plugging appropriate sample estimates into the definitions of H_{k1i} and H_{k2i}. It is easy to see that conditions (i) to (v) are fulfilled by independent standard normal multipliers G_i^{WB}, which corresponds to the original idea of Lin et al. (1993). Another option is, for example, independent centered unit Poisson multipliers, according to the suggestion of Beyersmann et al. (2013).

5 | PROOFS

Subsequently, we present the proofs of the lemmas and theorems from Sections 3 and 4.

Proof of Lemma 1. By the strong law of large numbers, we have

$$\begin{aligned} U_n(t) = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \Bigg(& \int_0^t \hat{S}\left(u - \middle| A = 1, \boldsymbol{Z_i}\right) \exp\left(\hat{\beta}_{1A} + \hat{\boldsymbol{\beta}}_{\boldsymbol{1Z}}^T \boldsymbol{Z_i}\right) \mathrm{d}\hat{\Lambda}_{01}(u) \\ & - \int_0^t S\left(u - \middle| A = 1, \boldsymbol{Z_i}\right) \exp\left(\beta_{01A} + \boldsymbol{\beta}_{\boldsymbol{01Z}}^T \boldsymbol{Z_i}\right) \mathrm{d}\Lambda_{01}(u) \Bigg) \\ & - \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \Bigg(\int_0^t \hat{S}\left(u - \middle| A = 0, \boldsymbol{Z_i}\right) \exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{1Z}}^T \boldsymbol{Z_i}\right) \mathrm{d}\hat{\Lambda}_{01}(u) \\ & - \int_0^t S\left(u - \middle| A = 0, \boldsymbol{Z_i}\right) \exp\left(\boldsymbol{\beta}_{\boldsymbol{01Z}}^T \boldsymbol{Z_i}\right) \mathrm{d}\Lambda_{01}(u) \Bigg) + o_p(1) \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\Bigg(\int_0^t \left(\hat{S}\left(u-\middle|A=1,\boldsymbol{Z_i}\right)-S\left(u-\middle|A=1,\boldsymbol{Z_i}\right)\right)\exp\left(\hat{\beta}_{1A}+\hat{\boldsymbol{\beta}}_{\boldsymbol{1Z}}^T\boldsymbol{Z_i}\right)\,\mathrm{d}\hat{\Lambda}_{01}(u)\\
&\quad+\int_0^t S\left(u-\middle|A=1,\boldsymbol{Z_i}\right)\,\mathrm{d}\left(\exp\left(\hat{\beta}_{1A}+\hat{\boldsymbol{\beta}}_{\boldsymbol{1Z}}^T\boldsymbol{Z_i}\right)\hat{\Lambda}_{01}(u)-\exp\left(\beta_{01A}+\boldsymbol{\beta}_{\boldsymbol{01Z}}^T\boldsymbol{Z_i}\right)\Lambda_{01}(u)\right)\Bigg)\\
&\quad-\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\Bigg(\int_0^t \left(\hat{S}\left(u-\middle|A=0,\boldsymbol{Z_i}\right)-S\left(u-\middle|A=0,\boldsymbol{Z_i}\right)\right)\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{1Z}}^T\boldsymbol{Z_i}\right)\,\mathrm{d}\hat{\Lambda}_{01}(u)\\
&\quad+\int_0^t S\left(u-\middle|A=0,\boldsymbol{Z_i}\right)\,\mathrm{d}\left(\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{1Z}}^T\boldsymbol{Z_i}\right)\hat{\Lambda}_{01}(u)-\exp\left(\boldsymbol{\beta}_{\boldsymbol{01Z}}^T\boldsymbol{Z_i}\right)\Lambda_{01}(u)\right)\Bigg)+o_p(1).
\end{aligned}
$$

Lin et al. (1994) showed that $\sqrt{n}\left(\hat{\Lambda}_k\left(t\middle|a,\boldsymbol{z}\right)-\Lambda_k\left(t\middle|a,\boldsymbol{z}\right)\right)=\tilde{W}_k\left(t\middle|a,\boldsymbol{z}\right)+o_p(1)$ for the martingale expression

$$
\begin{aligned}
\tilde{W}_k\left(t\middle|a,\boldsymbol{z}\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg(&\int_0^t\frac{\exp\left(\beta_{0kA}\cdot a+\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{z}\right)}{S^{(0)}(\boldsymbol{\beta_{0k}},u)}\,\mathrm{d}M_{ki}(u)\\
&+\left(\boldsymbol{h_k}\left(t\middle|a,\boldsymbol{z}\right)\right)^T\boldsymbol{\Sigma_k}^{-1}\int_0^\tau\left(\left(A_i,\boldsymbol{Z_i}^T\right)^T-\boldsymbol{E}(\boldsymbol{\beta_{0k}},u)\right)\,\mathrm{d}M_{ki}(u)\Bigg).
\end{aligned}
$$

Thus, exploiting the (uniform) consistency of $\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}$ and $\hat{\Lambda}_{01}$ (Kosorok, 2008, pp. 361–362; Tsiatis, 1981) and using a first-order Taylor approximation of $f: x\mapsto\exp(-x)$ around $x=\sum_{k=1}^K\exp\left(\beta_{0kA}\cdot a+\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{z}\right)\Lambda_{0k}(t)$ (which yields $\hat{S}\left(t-\middle|a,\boldsymbol{z}\right)-S\left(t-\middle|a,\boldsymbol{z}\right)=-\frac{1}{\sqrt{n}}\,S\left(t-\middle|a,\boldsymbol{z}\right)\sum_{k=1}^K\tilde{W}_k\left(t\middle|a,\boldsymbol{z}\right)+o_p(1)$), we find that

$$
\begin{aligned}
U_n(t)&=\frac{1}{n}\sum_{i=1}^{n}\Bigg(\int_0^t S\left(u-\middle|A=1,\boldsymbol{Z_i}\right)\,\mathrm{d}\tilde{W}_1\left(u\middle|A=1,\boldsymbol{Z_i}\right)\\
&\quad-\sum_{k=1}^{K}\int_0^t\tilde{W}_k\left(u\middle|A=1,\boldsymbol{Z_i}\right)\,\mathrm{d}F_1\left(u\middle|A=1,\boldsymbol{Z_i}\right)\Bigg)\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}\Bigg(\int_0^t S\left(u-\middle|A=0,\boldsymbol{Z_i}\right)\,\mathrm{d}\tilde{W}_1\left(u\middle|A=0,\boldsymbol{Z_i}\right)\\
&\quad-\sum_{k=1}^{K}\int_0^t\tilde{W}_k\left(u\middle|A=0,\boldsymbol{Z_i}\right)\,\mathrm{d}F_1\left(u\middle|A=0,\boldsymbol{Z_i}\right)\Bigg)+o_p(1)\\
&=\frac{1}{n}\sum_{i=1}^{n}\Bigg(\int_0^t S\left(u-\middle|A=1,\boldsymbol{Z_i}\right)\,\mathrm{d}\tilde{W}_1\left(u\middle|A=1,\boldsymbol{Z_i}\right)\\
&\quad-\sum_{k=1}^{K}\int_0^t\left(F_1\left(t\middle|A=1,\boldsymbol{Z_i}\right)-F_1\left(u\middle|A=1,\boldsymbol{Z_i}\right)\right)\,\mathrm{d}\tilde{W}_k\left(u\middle|A=1,\boldsymbol{Z_i}\right)\\
&\quad-\Bigg(\int_0^t S\left(u-\middle|A=0,\boldsymbol{Z_i}\right)\,\mathrm{d}\tilde{W}_1\left(u\middle|A=0,\boldsymbol{Z_i}\right)\\
&\quad-\sum_{k=1}^{K}\int_0^t\left(F_1\left(t\middle|A=0,\boldsymbol{Z_i}\right)-F_1\left(u\middle|A=0,\boldsymbol{Z_i}\right)\right)\,\mathrm{d}\tilde{W}_k\left(u\middle|A=1,\boldsymbol{Z_i}\right)\Bigg)\Bigg)+o_p(1).
\end{aligned}
$$

The last equivalence follows from integration by parts, since

$$\int_0^t \tilde{W}_k\left(u\middle|a,\boldsymbol{z}\right)\, \mathrm{d}F_1\left(u\middle|a,\boldsymbol{z}\right) = \tilde{W}_k\left(t\middle|a,\boldsymbol{z}\right)F_1\left(t\middle|a,\boldsymbol{z}\right) - \int_0^t F_1\left(u\middle|a,\boldsymbol{z}\right)\, \mathrm{d}\tilde{W}_k\left(u\middle|a,\boldsymbol{z}\right).$$

Finally, by inserting the definition of $\tilde{W}_k$ and reordering the terms, the result follows. □

Proof of Theorem 1. Lemma 1 implies that it is sufficient to consider the limiting distribution of $\tilde{U}_n$.

For distinct causes $k \neq l$, the counting processes N_{ki} and N_{li} cannot jump both, which is why the martingales $M_{ki}(t)$ and $M_{li}(t)$ are orthogonal. Moreover, $\forall k \in \{1, \ldots, K\}$,

$$\begin{aligned}
&\left\langle \sum_{i=1}^{n} \int_0^{\cdot} \frac{1}{\sqrt{n}\, S^{(0)}(\boldsymbol{\beta_{0k}}, u)}\, \mathrm{d}M_{ki}(u),\ \sum_{i=1}^{n} \int_0^{\cdot} \frac{1}{\sqrt{n}}\left(\left(A_i, \boldsymbol{Z_i}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}}, u)\right)\, \mathrm{d}M_{ki}(u) \right\rangle (t) \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta_{0k}}, u)}\left(\left(A_i, \boldsymbol{Z_i}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}}, u)\right) Y_i(u) \exp\left(\beta_{0kA} \cdot A_i + \boldsymbol{\beta_{0kZ}^T Z_i}\right)\, \mathrm{d}\Lambda_{0k}(u) \\
&= \int_0^t \frac{1}{S^{(0)}(\boldsymbol{\beta_{0k}}, u)}\left(\boldsymbol{S^{(1)}}(\boldsymbol{\beta_{0k}}, u) - \boldsymbol{E}(\boldsymbol{\beta_{0k}}, u) S^{(0)}(\boldsymbol{\beta_{0k}}, u)\right)\, \mathrm{d}\Lambda_{0k}(u) = \boldsymbol{0}.
\end{aligned}$$

This means that $\sum_{i=1}^n \int_0^t H_{k1i}(u,t)\, \mathrm{d}M_{ki}(u)$ and $\sum_{i=1}^n \int_0^t H_{k2i}(u,t)\, \mathrm{d}M_{ki}(u)$ are orthogonal as well. Andersen et al. (1993, pp. 498–501) furthermore showed that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \int_0^{\tau} \left(\left(A_i, \boldsymbol{Z_i}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}}, u)\right)\, \mathrm{d}M_{ki}(u) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_k}),$$

as n tends to infinity (where $\mathcal{N}$ symbolizes the normal distribution), and since $\boldsymbol{\varphi_1}$ and $\boldsymbol{\psi_{1k}}$ are deterministic functions for given covariates $\boldsymbol{Z_i}$, the second summand of $\tilde{U}_n$ is likewise asymptotically normal with mean zero.

It therefore only remains to consider the first summand. Note that $\forall k \in \{1, \ldots, K\}$, the processes $\tilde{H}_{k1}(u,t)$ are deterministic, continuous in u and bounded for fixed covariates $\boldsymbol{Z_i}$. In particular,

$$\begin{aligned}
\left|\tilde{H}_{k1}(u,t)\right| &\leq (\exp(\beta_{0kA}) + 1) \max_{1 \leq i \leq n} \exp\left(\boldsymbol{\beta_{0kZ}^T Z_i}\right), \\
\left|\left(\tilde{H}_{k1}(u,t)\right)^2\right| &\leq (\exp(2\beta_{0kA}) + 2\exp(\beta_{0kA}) + 1) \max_{1 \leq i,j \leq n} \exp\left(\boldsymbol{\beta_{0kZ}^T}(\boldsymbol{Z_i} + \boldsymbol{Z_j})\right),
\end{aligned}$$

for $u \leq t$. The strong law of large numbers further suggests that $S^{(0)}(\boldsymbol{\beta_k}, t)$ converges to $s^{(0)}(\boldsymbol{\beta_k}, t)$ almost surely for any $t \in [0, \tau]$, $\boldsymbol{\beta_k} \in \mathbb{R}^{p_k+1}$. If we suppose that $P(Y_i(\tau) > 0) > 0\ \forall i \in \{1, \ldots, n\}$ (or also some less stringent constraints, see Fleming & Harrington, 2005, Sec. 8.4), this convergence is uniform on $\mathcal{B}_k \times [0, \tau]$, where $\mathcal{B}_k$ is a neighbourhood of $\boldsymbol{\beta_{0k}}$. Besides, $s^{(0)}$ is bounded away from zero on $\mathcal{B}_k \times [0, \tau]$. The conditions of Rebolledo's martingale central limit theorem (Andersen et al., 1993, Thm. II.5.1) are thus fulfilled, and we conclude that $\tilde{U}_n$ converges weakly to a zero-mean Gaussian process on $\mathcal{D}[0, \tau]$.

For the covariance function $\tilde{\xi}$, one finds that

$$
\begin{aligned}
\tilde{\xi}(t_1,t_2) = \sum_{k=1}^{K}\Bigg(& \sum_{i=1}^{n}\int_0^{t_1\wedge t_2} \frac{\tilde{H}_{k1}(u,t_1)\,\tilde{H}_{k1}(u,t_2)}{n\left(S^{(0)}(\boldsymbol{\beta_{0k}},u)\right)^2}\; Y_i(u)\exp\left(\beta_{0kA}\cdot A_i + \boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z_i}\right)\,\mathrm{d}\Lambda_{0k}(u) \\
& + \sum_{i=1}^{n}\int_0^{\tau}\frac{1}{n}\left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_1)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}},u)\right) \\
& \times\left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_2)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}},u)\right)\cdot Y_i(u)\exp\left(\beta_{0kA}\cdot A_i + \boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z_i}\right)\,\mathrm{d}\Lambda_{0k}(u)\Bigg) \\
= \sum_{k=1}^{K}\Bigg(& \int_0^{t_1\wedge t_2}\frac{\tilde{H}_{k1}(u,t_1)\,\tilde{H}_{k1}(u,t_2)}{S^{(0)}(\boldsymbol{\beta_{0k}},u)}\,\mathrm{d}\Lambda_{0k}(u) \\
& + \left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_1)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\Bigg(\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}},u)\right)\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\boldsymbol{\beta_{0k}},u)\right)^T \\
& \cdot Y_i(u)\exp\left(\beta_{0kA}\cdot A_i + \boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z_i}\right)\,\mathrm{d}\Lambda_{0k}(u)\Bigg)\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\right)^T\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_2)\Bigg) \\
= \sum_{k=1}^{K}\Bigg(& \int_0^{t_1\wedge t_2}\frac{\tilde{H}_{k1}(u,t_1)\,\tilde{H}_{k1}(u,t_2)}{S^{(0)}(\boldsymbol{\beta_{0k}},u)}\,\mathrm{d}\Lambda_{0k}(u) \\
& + \left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_1)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\int_0^{\tau}\left(\boldsymbol{S}^{(2)}(\boldsymbol{\beta_{0k}},u) - \boldsymbol{S}^{(1)}(\boldsymbol{\beta_{0k}},u)\left(\boldsymbol{E}(\boldsymbol{\beta_{0k}},u)\right)^T\right)\mathrm{d}\Lambda_{0k}(u)\right)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\,\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_2)\Bigg) \\
\xrightarrow[n\to\infty]{} & \sum_{k=1}^{K}\left(\int_0^{t_1\wedge t_2}\frac{\tilde{H}_{k1}(u,t_1)\,\tilde{H}_{k1}(u,t_2)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}\,\mathrm{d}\Lambda_{0k}(u) + \left(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_1)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\,\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_2)\right) \\
= \xi(t_1,t_2), &
\end{aligned}
$$

where the convergence in the last step follows by the strong law of large numbers and the continuous mapping theorem. □

Proof of Theorem 2 ((Outline)). The superscript "∗" is used here and in the following to indicate bootstrapped quantities.

Suppose that the given data are obtained on the probability space $(\Omega,\mathcal{A},P)$. We first note that the general martingale arguments apply conditionally on $\omega\in\Omega$ for almost all ω (cf. Akritas, 1986). Let $\tau^* = \max_{1\le i\le n}\{(T\wedge C)_i^*\}$ be the last observed event time in the bootstrap sample. The estimators $\hat{\Lambda}_{0k}$ and $\hat{\Lambda}_k$ calculated in the original data sample are now the true (discontinuous) cumulative baseline hazard and cumulative hazard in the bootstrap sample, respectively. Moreover, $s^{(r),*}(\boldsymbol{\beta_k},u) = S^{(r)}(\boldsymbol{\beta_k},u)$ as well as $\boldsymbol{e}^*(\boldsymbol{\beta_k},u) = \boldsymbol{E}(\boldsymbol{\beta_k},u)$, which is easy to see if the bootstrap sample is represented with multinomial weights assigned to the original sample, for example,

$$
S^{(0),*}(\boldsymbol{\beta_k},u) = \frac{1}{n}\sum_{i=1}^{n} w_i\; Y_i(u)\exp\left(\beta_{kA}A_i + \boldsymbol{\beta}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right),
$$

for $\boldsymbol{w}\sim\mathcal{M}ult\left(n,\left(1/n,\dots,1/n\right)^T\right)$. Thus, $\boldsymbol{\Sigma}_{\boldsymbol{k}}^* = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}$, $\boldsymbol{h}_{\boldsymbol{k}}^* = \hat{\boldsymbol{h}}_{\boldsymbol{k}}$, $\boldsymbol{\varphi}_{\boldsymbol{1}}^* = \hat{\boldsymbol{\varphi}}_{\boldsymbol{1}}$, $\boldsymbol{\psi}_{\boldsymbol{1k}}^* = \hat{\boldsymbol{\psi}}_{\boldsymbol{1k}}$ and

$$
M_{ki}^*(t) = w_i\left(N_{ki}(t) - \int_0^t Y_i(u)\,\mathrm{d}\hat{\Lambda}_k\left(u\middle|A_i,\boldsymbol{Z_i}\right)\right),
$$

for $i \in \{1,\dots,n\}$. Note that a discrete-time setting is considered here! We can now infer that $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{k}} \xrightarrow{P} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ and $\hat{\Lambda}^*_{0k} \xrightarrow{a.s.} \hat{\Lambda}_{0k}$ on $[0,\tau^*]$ as $n \to \infty$ by the considerations of Prentice and Kalbfleisch (2003). Also,

$$\begin{aligned}
&\sqrt{n}\Big(\hat{\Lambda}^*_k\Big(t\Big|a,\boldsymbol{z}\Big) - \hat{\Lambda}_k\Big(t\Big|a,\boldsymbol{z}\Big)\Big)\\
&= \frac{1}{\sqrt{n}}\int_0^t \frac{\exp\left(\hat{\beta}_{kA}\cdot a + \hat{\boldsymbol{\beta}}^T_{\boldsymbol{kZ}}\boldsymbol{z}\right)}{S^{(0),*}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\, M^*_k(\mathrm{d}u)\\
&\quad + \frac{1}{\sqrt{n}}\Big(\hat{\boldsymbol{h}}_{\boldsymbol{k}}\Big(t\Big|a,\boldsymbol{z}\Big)\Big)^T\hat{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{k}}\left(\sum_{i=1}^n\int_0^{\tau^*}\Big(\big(A_i,\boldsymbol{Z}^T_{\boldsymbol{i}}\big)^T - \boldsymbol{E}^*(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)\, M^*_{ki}(\mathrm{d}u)\right) + o_p(1),
\end{aligned}$$

which can be concluded by the reasoning of Andersen et al. (1993, proof of Thm. VII.2.3). These results provide the basis for proceeding in the same way as we did in the proof of Lemma 1. It follows that $U^*_n(t) = \tilde{U}^*_n(t) + o_p(1)$, with

$$\tilde{U}^*_n(t) = \sum_{k=1}^K\sum_{i=1}^n\left(\int_0^t H^*_{k1i}(u,t)\, M^*_{ki}(\mathrm{d}u) + \int_0^{\tau^*} H^*_{k2i}(u,t)\, M^*_{ki}(\mathrm{d}u)\right),$$

applying the definitions of H_{k1i} and H_{k2i} to the bootstrap sample.

Subsequently, we use similar arguments as in the proof of Theorem 1. One finds that

$$\begin{aligned}
&\left\langle \sum_{i=1}^n\int_0^{\cdot}\frac{1}{\sqrt{n}\,S^{(0),*}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\, M^*_{ki}(\mathrm{d}u),\ \sum_{i=1}^n\int_0^{\cdot}\frac{1}{\sqrt{n}}\Big(\big(A_i,\boldsymbol{Z}^T_{\boldsymbol{i}}\big)^T - \boldsymbol{E}^*(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)\, M^*_{ki}(\mathrm{d}u)\right\rangle(t)\\
&= \frac{1}{n}\sum_{i=1}^n\int_0^t\frac{1}{S^{(0),*}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\Big(\big(A_i,\boldsymbol{Z}^T_{\boldsymbol{i}}\big)^T - \boldsymbol{E}^*(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)\\
&\quad\times w_i\, Y_i(u)\Big(1-\hat{\Lambda}_k\Big(\Delta u\Big|A_i,\boldsymbol{Z}_{\boldsymbol{i}}\Big)\Big)\,\hat{\Lambda}_k\Big(\mathrm{d}u\Big|A_i,\boldsymbol{Z}_{\boldsymbol{i}}\Big)\\
&\to \int_0^t\frac{1}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\Big(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)S^{(0)}_2(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u) - \boldsymbol{S}^{(\boldsymbol{1})}_{\boldsymbol{2}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)\hat{\Lambda}_{0k}(\Delta u)\,\hat{\Lambda}_{0k}(\mathrm{d}u),
\end{aligned}$$

with $S^{(r)}_2(\boldsymbol{\beta}_{\boldsymbol{k}},u) = \frac{1}{n}\sum_{i=1}^n Y_i(t)\exp\left(2\beta_{kA}A_i + 2\boldsymbol{\beta}^T_{\boldsymbol{kZ}}\boldsymbol{Z}_{\boldsymbol{i}}\right)\Big(\big(A_i,\boldsymbol{Z}^T_{\boldsymbol{i}}\big)^T\Big)^{\otimes r}$, $r \in \{0,1,2\}$. Also,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^{\tau^*}\Big(\big(A_i,\boldsymbol{Z}^T_{\boldsymbol{i}}\big)^T - \boldsymbol{E}^*(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)\, M^*_{ki}(\mathrm{d}u) \xrightarrow{\mathscr{D}} \mathcal{N}\Big(\boldsymbol{0},\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k},\boldsymbol{2}}\Big),$$

with

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k},\boldsymbol{2}} = \int_0^{\tau}&\Big(\boldsymbol{S}^{(\boldsymbol{2})}_{\boldsymbol{2}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u) - \boldsymbol{S}^{(\boldsymbol{1})}_{\boldsymbol{2}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\big(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\big)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big(\boldsymbol{S}^{(\boldsymbol{1})}_{\boldsymbol{2}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)^T\\
&+\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\big(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\big)^T S^{(0)}_2(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\Big)\\
&\cdot\hat{\Lambda}_{0k}(\Delta u)\,\hat{\Lambda}_{0k}(\mathrm{d}u),
\end{aligned}$$

(cf. Prentice & Kalbfleisch, 2003), and lastly,

$$\sum_{i=1}^{n}\int_0^t \frac{\tilde{H}^*_{k1}(u,t)}{\sqrt{n}\; S^{(0),*}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\; M^*_{ki}(\mathrm{d}u),$$

converges weakly to a zero-mean Gaussian process with covariance function

$$\begin{aligned}\xi_1^*(t_1,t_2) = &\sum_{k=1}^{K}\int_0^{t_1\wedge t_2}\frac{\tilde{H}_{k1}(u,t_1)\;\tilde{H}_{k1}(u,t_2)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\;\hat{\Lambda}_{0k}(\mathrm{d}u)\\ &-\sum_{k=1}^{K}\int_0^{t_1\wedge t_2}\frac{\tilde{H}_{k1}(u,t_1)\;\tilde{H}_{k1}(u,t_2)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}\;S_2^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\hat{\Lambda}_{0k}(\Delta u)\;\hat{\Lambda}_{0k}(\mathrm{d}u),\end{aligned}$$

as $n\to\infty$.

Since we assumed that there are no ties in the original sample,

$$\begin{aligned}S_2^{(r)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\hat{\Lambda}_{0k}(\Delta u)\;\hat{\Lambda}_{0k}(\mathrm{d}u) &= \frac{1}{n}\sum_{i=1}^{n}Y_i(u)\exp\left(2\hat{\beta}_{kA}A_i + 2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right)\left(\left(A_i,\boldsymbol{Z_i}^T\right)^T\right)^{\otimes r}\\ &\quad\times\frac{(\Delta N_k(u))^2}{\left(\sum_{i=1}^{n}Y_i(u)\exp\left(\hat{\beta}_{kA}A_i+\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right)\right)^2}\\ &\le\frac{\max\limits_{1\le i\le n:\;Y_i(u)=1}\left\{\exp\left(2\hat{\beta}_{kA}A_i+2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right)\left(\left(A_i,\boldsymbol{Z_i}^T\right)^T\right)^{\otimes r}\right\}}{(Y(u))^2\min\limits_{1\le i\le n:\;Y_i(u)=1}\left\{\exp\left(2\hat{\beta}_{kA}A_i+2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right)\right\}}.\end{aligned}$$

Because of the boundedness of the covariates, all the terms involving $S_2^{(r)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\hat{\Lambda}_{0k}(\Delta u)\;\hat{\Lambda}_{0k}(\mathrm{d}u)$, $r\in\{0,1,2\}$, vanish as $n\to\infty$, and the proof is complete. □

Before we can verify Theorem 3, several interim results are needed. The proofs of the following lemmas can be found in the appendix; the ideas are based on Beyersmann et al. (2013) and Dobler et al. (2017).

Consider the triangular arrays $\boldsymbol{X}_{\boldsymbol{n,i}}^{(\boldsymbol{k})}=\left(X_{n,i}^{(k)}(t_1),\dots,X_{n,i}^{(k)}(t_l)\right)^T$, $i\in\{1,\dots,n\}$, $k\in\{1,\dots,K\}$, defined on the probability space $(\Omega_1,\mathcal{A}_1,P_1)$, for $0\le t_1\le\ldots\le t_l\le\tau$, $l\in\mathbb{N}$, with

$$X_{n,i}^{(k)}(t)=\int_0^t\hat{\hat{H}}_{k1}(u,t)\;\frac{\mathrm{d}N_{ki}(u)}{\sqrt{n}\;S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}+\int_0^\tau\frac{1}{\sqrt{n}}\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z_i}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\mathrm{d}N_{ki}(u),$$

constituting the subject- and cause-specific summands of $\tilde{U}_n(t)$ (except that $\mathrm{d}M_{ki}$ is replaced by $\mathrm{d}N_{ki}$), as well as the plug-in estimators $\hat{\hat{H}}_{k1}$, $\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}$ and

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left(\frac{\boldsymbol{S}^{(\boldsymbol{2})}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}-\left(\frac{\boldsymbol{S}^{(\boldsymbol{1})}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\right)^{\otimes 2}\right)\mathrm{d}N_{ki}(u).$$

Consequently, $\hat{U}_n(t)=\sum_{k=1}^{K}\sum_{i=1}^{n}G_i\,X_{n,i}^{(k)}(t)$, with multipliers G_i defined on $(\Omega_2,\mathcal{A}_2,P_2)$. (We generally consider the product probability space $(\Omega,\mathcal{A},P)=(\Omega_1\times\Omega_2,\mathcal{A}_1\otimes\mathcal{A}_2,P_1\times P_2)$.)

Lemma 2. *The triangular arrays* $\boldsymbol{X}_{\boldsymbol{n,i}}^{(\boldsymbol{k})}$ *satisfy the following conditions for each* $k \in \{1,\dots,K\}$:

(i) $\max_{1\le i\le n} \left\| \boldsymbol{X}_{\boldsymbol{n,i}}^{(\boldsymbol{k})} \right\| \xrightarrow{P} 0$ *(where* $\|\cdot\|$ *denotes the Euclidian norm)*;

(ii) $\sum_{i=1}^{n} \boldsymbol{X}_{\boldsymbol{n,i}}^{(\boldsymbol{k})} \left(\boldsymbol{X}_{\boldsymbol{n,i}}^{(\boldsymbol{k})}\right)^T \xrightarrow{P} \left(\xi^{(k)}(t_r,t_s)\right)_{1\le r,s\le l}$.

Lemma 3. *For time points* $0 \le t_r \le t_s \le \tau$ *and* $k \in \{1,\dots,K\}$,

$$\max_{1\le i\le n} \left| X_{n,i}^{(k)}(t_s) - X_{n,i}^{(k)}(t_r) \right| \in O_p\left(n^{-1/2}\right),$$

where $O_p(a_n)$ *denotes asymptotic boundedness by* a_n *in probability.*

Note that the bound in Lemma 3 is independent of the time points t_r and t_s!

Lemma 4. *For time points* $0 \le t_q \le t_r \le t_s \le \tau$*, causes* $k \in \{1,\dots,K\}$ *and the function*

$$\begin{aligned}
L_n^{(k)}(t) = {} & \frac{1}{n}\sum_{j=1}^{n} \left(\exp\left(2\hat{\beta}_{kA}\right) + 2\exp\left(\hat{\beta}_{kA}+1\right)\right) \exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{Z_j}\right) \int_0^t \frac{\mathrm{d}N_k(u)}{n\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2} \\
& + \frac{1}{n}\sum_{j=1}^{n} \left(\left(\hat{F}_1\left(t \middle| A=1,\boldsymbol{Z_j}\right)\right)^2 \exp\left(2\hat{\beta}_{kA}\right) + \left(\hat{F}_1\left(t \middle| A=0,\boldsymbol{Z_j}\right)\right)^2 \right) \\
& \times \exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{Z_j}\right) \int_0^\tau \frac{\mathrm{d}N_k(u)}{n\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2} \\
& + \frac{1}{n}\sum_{j=1}^{n} \exp\left(2\hat{\beta}_{kA} + 2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{Z_j}\right) \\
& \cdot \int_0^t \left(\left(1,\boldsymbol{Z_j^T}\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\right) \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1} \\
& \times \left(\left(1,\boldsymbol{Z_j^T}\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right) \frac{\mathrm{d}N_k(u)}{n\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2} \\
& + \frac{1}{n}\sum_{j=1}^{n} \exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{Z_i}\right) \\
& \cdot \int_0^t \left(\left(0,\boldsymbol{Z_i^T}\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\right) \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1} \\
& \times \left(\left(0,\boldsymbol{Z_i^T}\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right) \frac{\mathrm{d}N_k(u)}{n\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2} \\
& + \frac{1}{n}\sum_{j=1}^{n} \left(\hat{F}_1\left(t \middle| A=1,\boldsymbol{Z_j}\right)\right)^2 \exp\left(2\hat{\beta}_{kA} + 2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T \boldsymbol{Z_j}\right) \\
& \cdot \int_0^\tau \left(\left(1,\boldsymbol{Z_j^T}\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\right) \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1} \\
& \times \left(\left(1,\boldsymbol{Z_j^T}\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right) \frac{\mathrm{d}N_k(u)}{n\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}
\end{aligned}$$

$$+\frac{1}{n}\sum_{j=1}^{n}\left(\hat{F}_1\left(t\middle|A=0,\boldsymbol{Z_i}\right)\right)^2\exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right)$$
$$\cdot\int_0^\tau\left(\left(0,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}$$
$$\times\left(\left(0,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\frac{\mathrm{d}N_k(u)}{n\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}.$$

with $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}=\int_0^\tau\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T\mathrm{d}N_{ki}(u)$, *the following inequality holds in probability provided that the conditions in Theorem 3 are fulfilled*:

$$\mathbb{E}\left(\left(\sum_{i=1}^{n}G_i\,X_{n,i}^{(k)}(t_r)-\sum_{i=1}^{n}G_i\,X_{n,i}^{(k)}(t_q)\right)^2\left(\sum_{i=1}^{n}G_i\,X_{n,i}^{(k)}(t_s)-\sum_{i=1}^{n}G_i\,X_{n,i}^{(k)}(t_r)\right)^2\middle|\mathscr{F}_\tau\right)$$
$$\leq\left(L_n^{(k)}(t_s)-L_n^{(k)}(t_q)\right)^{3/2}\cdot O_p(1).$$

Proof of Theorem 3. Considering $\hat{U}_n^{(k)}(\cdot)=\sum_{i=1}^{n}G_i\,X_{n,i}^{(k)}(\cdot)$, the conditions of Lemma 1 from the supplementary material of Dobler et al. (2017) are fulfilled $\forall k\in\{1,\ldots,K\}$ due to Lemma 2 and the assumptions w.r.t. the multipliers G_i. It follows that the finite-dimensional distributions of $\hat{U}_n^{(k)}$ converge weakly to zero-mean Gaussian processes with covariance functions $\xi^{(k)}$, respectively, in probability (conditional on $\mathscr{F}_\tau$).

Since $\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}$ converges to $\boldsymbol{\Sigma_k}$ (see proof of Lemma 2), the function $L_n^{(k)}$ from Lemma 4 converges uniformly to

$$l^{(k)}(t)=(\exp(2\beta_{0kA})+2\exp(\beta_{0kA}+1))\mathbb{E}_{\boldsymbol{Z}}\left(\exp\left(2\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z}\right)\right)\int_0^t\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}$$
$$+\mathbb{E}_{\boldsymbol{Z}}\left(\left(\left(F_1\left(t\middle|A=1,\boldsymbol{Z}\right)\right)^2\exp(2\beta_{0kA})+\left(F_1\left(t\middle|A=0,\boldsymbol{Z}\right)\right)^2\right)\right.$$
$$\left.\times\exp\left(2\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z}\right)\right)\int_0^\tau\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}$$
$$+\mathbb{E}_{\boldsymbol{Z}}\left(\exp\left(2\beta_{0kA}+2\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z}\right)\int_0^t\left(\left(1,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\right.$$
$$\left.\times\left(\left(1,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\right)\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}\right)$$
$$+\mathbb{E}_{\boldsymbol{Z}}\left(\exp\left(2\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z}\right)\int_0^t\left(\left(0,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\right.$$
$$\left.\times\left(\left(0,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\right)\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}\right)$$
$$+\mathbb{E}_{\boldsymbol{Z}}\left(\left(F_1\left(t\middle|A=1,\boldsymbol{Z}\right)\right)^2\exp\left(2\beta_{0kA}+2\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z}\right)\right.$$
$$\left.\cdot\int_0^\tau\left(\left(1,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\left(1,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\right)\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}\right)$$

$$+\mathbb{E}_{\boldsymbol{Z}}\left(\left(F_1\left(t\middle|A=0,\boldsymbol{Z}\right)\right)^2 \exp\left(2\boldsymbol{\beta}_{\boldsymbol{0kZ}}^T\boldsymbol{Z}\right)\right.$$
$$\left.\cdot\int_0^{\tau}\left(\left(0,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta}_{\boldsymbol{0k}},u)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\left(0,\boldsymbol{Z}^T\right)^T-\boldsymbol{e}(\boldsymbol{\beta}_{\boldsymbol{0k}},u)\right)\frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta}_{\boldsymbol{0k}},u)}\right),$$

on $[0,\tau)$ as a consequence of the martingale central limit theorem. The conditional tightness of $\hat{U}_n^{(k)}$ can now be shown along the lines of the proof of Theorem 3.1 in Dobler and Pauly (2014). We apply the subsequence principle for convergence in probability (cf. Beyersmann et al., 2013): For every subsequence, there is another subsequence such that for almost every (fixed) $\omega \in \Omega_1 \times \Omega_2$, we find $n_0 \in \mathbb{N}$, a constant $\gamma > 0$ and a sequence of nondecreasing, continuous functions $l_n^{(k)}$ that converges uniformly to $l^{(k)}$, such that

$$\mathbb{E}\left(\left(\sum_{i=1}^n G_i\, X_{n,i}^{(k)}(t_r)-\sum_{i=1}^n G_i\, X_{n,i}^{(k)}(t_q)\right)^2\left(\sum_{i=1}^n G_i\, X_{n,i}^{(k)}(t_s)-\sum_{i=1}^n G_i\, X_{n,i}^{(k)}(t_r)\right)^2\middle|\mathcal{F}_\tau\right)$$
$$\leq \gamma\left(l_n^{(k)}(t_s)-l_n^{(k)}(t_q)\right)^{3/2},$$

if $n \geq n_0$. (Here, n_0 and γ do not depend on $0 \leq t_q \leq t_r \leq t_s \leq \tau$.) The conditional tightness follows by extending Theorem 13.5 in Billingsley (1999) pointwise along subsequences (cf. Dobler & Pauly, 2014). Eventually, this proves the conditional convergence in distribution of $\hat{U}_n^{(k)}$ in probability for each $k \in \{1,\ldots,K\}$.

The assertion of Theorem 3 follows by noting that the processes $\hat{U}_n^{(k)}$ and $\hat{U}_n^{(k\prime)}$ are independent for $k \neq k'$ given the data because we consider competing events, that is, $\mathrm{d}N_{ki}$ and $\mathrm{d}N_{k'i}$ cannot jump both. □

6 | COMPARISON OF THE RESAMPLING METHODS

To provide better insight into the properties of the discussed resampling methods, the outcomes of extensive simulations investigating their performance are summarized below. Besides, we present the results of an application to real data.

6.1 | Performance in simulations

In the simulation study described in Rühl and Friedrich (2024), competing risks data were generated according to the same setup as used for the empirical studies presented by Ozenne et al. (2020): The authors considered 12 independent covariates with distinct effects on the probability of treatment as well as on the event of interest, a competing event, and censoring. A data generation approach based on cause-specific hazards was employed to model the event times. In the next step, Efron's bootstrap, the influence function approach, and the wild bootstrap (with standard normal, centred Poisson, and weird bootstrap multipliers) were applied to determine confidence intervals and time-simultaneous confidence bands, respectively.

Rühl and Friedrich (2024) investigated different scenarios, including settings with no, light or heavy censoring, low or high probability of treatment, and differing variance of the covariates. Sample sizes ranged between 50 and 300 and in addition, the extent of the treatment effect varied. Each scenario was simulated 5000 times and the performance of the three resampling methods was compared considering the 95% coverage probabilities as well as the mean widths of the confidence intervals and bands.

In general, the confidence intervals computed by the wild bootstrap achieved coverage probabilities that were the closest to the nominal level of 95%, in particular at later time points when a sufficient amount of events had been observed (regardless of the choice of the multiplier). Efron's bootstrap was more conservative, yielding coverages above those produced by the wild bootstrap, whereas the intervals based on the influence function attained the lowest coverage levels. Exceptions occurred, however, if the event of interest was observed only rarely, due to a prevalence of the competing event. In that case, all methods produced conservative confidence intervals, such that the influence function approach became (slightly) more accurate than the remaining approaches. The same applied if the sample size was very small (i.e., below 75–100) or covariates varied strongly.

The simulations revealed similar outcomes with respect to the time-simultaneous confidence bands, in the sense that the classical bootstrap and the influence function approach yielded more conservative and more liberal bands than the wild bootstrap, respectively. However, there was an increased number of settings where the coverage probability of the bands obtained by the wild bootstrap was less accurate compared to Efron's bootstrap.

The widths of the confidence intervals and bands corresponding to the influence function were overall the smallest. With increasing sample sizes, the differences between the methods became irrelevant, though.

Last but not least, it should be mentioned that Efron's bootstrap takes considerably more computation time than the influence function approach and the wild bootstrap. This might be relevant in practical analyses, in particular when sample sizes are large.

For more details on the simulations, readers are referred to Rühl and Friedrich (2024).

6.2 | Real data application

The application of the three resampling methods is further exemplified by means of data collected from the OAI (Nevitt et al., 2006). In a prospective cohort study, adults who suffer from or are at risk of symptomatic femoral-tibial knee osteoarthritis were enrolled at five clinical sites located in the United States, and between 2004 and 2014, the progression of their osteoarthritis was monitored. At the 48-month visit, researchers additionally obtained minute-by-minute accelerometer counts to record the participants' physical activity.

The goal of our analysis is to examine the effect of attaining the aerobic guideline of the U.S. Department of Health and Human Services (DHHS) (2008) for adults with arthritis (i.e., spending 150 min per week with moderately-to-vigorously intense activity that leads to at least 2020 activity counts per minute) on the time to knee replacement. The latter is determined by self-report or medical records, and death (as obtained from documentation) is considered as competing event. We only take into account individuals who contributed valid accelerometry data for at least four consecutive days (with activity counts being measured for 10 or more hours per day, as well as weekly activity being estimated on the basis of the available wear days if necessary). Eligible subjects also need to have a Kellgren–Lawrence grade of 3 or 4 in at least one knee (as agreed upon by two readers assessing the respective x-rays), while not having had any knee replacement until the 48-month visit (cf. Master et al., 2021). Besides, only participants with complete

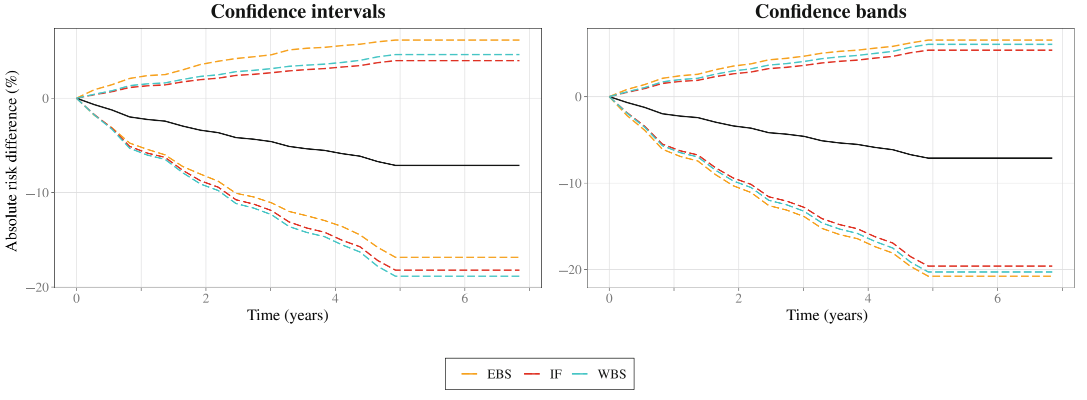

FIGURE 1 Confidence intervals and bands for the average effect of attaining the US Department of Health and Human Services (DHHS) activity guideline on the risk of knee replacement.

data on the covariates presented in Table A1 are investigated. Our analysis ultimately comprises 461 subjects.

The estimated average effect of attaining the DHHS activity guideline on the risk of knee replacement is illustrated in Figure 1: Our analysis implies that if every participant had achieved the recommended amount of physical activity, the 6-year risk of knee replacement would have been reduced by 7.11% as opposed to a setting where no one had attained the guideline. Confidence intervals and time-simultaneous confidence bands for the average effect are calculated using the resampling methods presented earlier. As it can be seen, the classical bootstrap yields confidence intervals with somewhat greater limits. Other than that, the influence function approach leads to the narrowest confidence regions and the wild bootstrap (with standard normal multipliers) produces slightly wider limits, which is in line with the outcomes of the simulation study discussed in the previous section.

7 | DISCUSSION

In this manuscript, we consider time-to-event data subject to competing risks and defined the ATE as the difference between the *t*-year absolute risks for the event of interest. We derived the asymptotic distribution of the g-formula estimator and on the basis thereof, examined three resampling methods that are useful for statistical inference. These include Efron's nonparametric bootstrap, an approach based on the influence function and the wild bootstrap. Simulations (Rühl & Friedrich, 2024) as well as an applied data analysis showed that confidence regions derived by the classical bootstrap generally were wider, whereas the influence function approach results in the least conservative intervals and bands. The wild bootstrap mostly ranged in between and therefore suggests itself as a reasonable choice, unless the event of interest is rarely observed. In that case, one might prefer the influence function. It is further worth noting that—in particular, for large samples—the computation time of the classical bootstrap significantly exceeds that of the two remaining approaches.

Beside quantifying its uncertainty, estimating the ATE also presents a number of challenges. While we based our estimation on cause-specific Cox models, our estimand of interest was the causal risk difference, i.e. a contrast of the cumulative incidence functions. Thus, we do not interpret the hazard ratio in a causal way, see Hernán (2010), Aalen et al. (2015), Martinussen and Vansteelandt (2013) for detailed discussions on the drawbacks of the hazard ratio in a causal context. As Martinussen and Stensrud (2023) point out, this estimand only captures the total effect of the treatment on the event of interest, while a distinction in terms of direct and indirect effects is not possible, see also Young et al. (2020). Alternative approaches (which both rely on untestable assumptions, though) are discussed by Rubin (2006) and Stensrud et al. (2022), including principal stratification and separable effects. Martinussen and Stensrud (2023) propose an estimator for the latter based on the efficient influence function and use the nonparametric bootstrap to estimate its variance. As the classical bootstrap has been shown to perform insufficiently in certain situations (Friedrich et al., 2017; Nießl et al., 2023; Rühl et al., 2023; Singh, 1981), extensions of the wild bootstrap to this situation merit further research.

Another aspect concerns modeling of the association between the covariates and the outcome. In our work, we focused on Cox proportional hazards models, but did not go into aspects such as variable selection or model misspecification. The latter is covered to some extent by Ozenne et al. (2020) considering the classical bootstrap. Recently, Vansteelandt et al. (2022) proposed an approach that allows for more flexible modeling of the association between covariates and an outcome. Integration of this so-called assumption-lean Cox regression into our resampling framework is part of future research. Alternatively, other regression models such as Aalen's additive hazards model (Aalen, 1989) or a Cox–Aalen model (Scheike & Zhang, 2002) might be used, however, the proofs presented here need to be adapted accordingly. Exploiting the properties of the martingale residuals underlying these models will be helpful and, in addition, facilitate the integration of, for example, left-truncation.

ACKNOWLEDGMENTS

Support by the DFG (Grant FR 4121/2-1) is gratefully acknowledged. The authors further thank Dennis Dobler for his constructive ideas and appreciate the insightful comments of two anonymous reviewers, which helped to improve and clarify this manuscript. Data and/or research tools used in the preparation of this manuscript were obtained and analysed from the controlled access datasets distributed from the OAI, a data repository housed within the NIMH Data Archive (NDA). OAI is a collaborative informatics system created by the National Institute of Mental Health and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to provide a worldwide resource to quicken the pace of biomarker identification, scientific investigation and OA drug development. Dataset identifier: 10.15154/ja67-pr39. Open Access funding enabled and organized by Projekt DEAL.

ORCID

Jasmin Rühl https://orcid.org/0000-0002-1721-0640

REFERENCES

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, *8*(8), 907–925.

Aalen, O. O., Cook, R. J., & Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*, *21*(4), 579–593.

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of the American Statistical Association*, *81*(396), 1032–1038.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes* (1st ed.). Springer.

Benichou, J., & Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, *46*(3), 813–826.

Beyersmann, J., Di Termini, S., & Pauly, M. (2013). Weak convergence of the wild bootstrap for the Aalen-Johansen estimator of the cumulative incidence function of a competing risk. *Scandinavian Journal of Statistics*, *40*(3), 387–402.

Billingsley, P. (1999). *Convergence of probability measures* (2nd ed.). John Wiley & Sons.

Breslow, N. E. (1972). Contribution to the discussion of paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *34*(2), 216–217.

Cheng, S. C., Fine, J. P., & Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, *54*(1), 219–228.

Dobler, D., Beyersmann, J., & Pauly, M. (2017). Non-strange weird resampling for complex survival data. *Biometrika*, *104*(3), 699–711.

Dobler, D., & Pauly, M. (2014). Bootstrapping Aalen-Johansen processes for competing risks: Handicaps, solutions, and limitations. *Electronic Journal of Statistics*, *8*(2), 2779–2803.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, *76*(374), 312–319.

Fleming, T. R., & Harrington, D. P. (2005). *Counting processes and survival analysis* (2nd ed.). John Wiley & Sons.

Friedrich, S., Brunner, E., & Pauly, M. (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis*, *153*, 255–265.

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology*, *21*(1), 13–15.

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman&Hall/CRC.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference* (1st ed.). Springer.

Lin, D. Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics and Medicine*, *16*(8), 901–910.

Lin, D. Y., Fleming, L. J., & Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika*, *81*(1), 73–81.

Lin, D. Y., Wei, L. J., & Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, *80*(3), 557–572.

Martinussen, T., & Stensrud, M. J. (2023). Estimation of separable direct and indirect effects in continuous time. *Biometrics*, *79*(1), 127–139.

Martinussen, T., & Vansteelandt, S. (2013). On collapsibility and confounding bias in cox and Aalen regression models. *Lifetime Data Analysis*, *19*(3), 279–296.

Master, H., Thoma, L. M., Neogi, T., Dunlop, D. D., LaValley, M., Christiansen, M. B., & White, D. K. (2021). Daily walking and the risk of knee replacement over five years among adults with advanced knee osteoarthritis in the United States. *Archives of Physical Medicine and Rehabilitation*, *102*(10), 1888–1894.

Neumann, A., & Billionnet, C. (2016). Covariate adjustment of cumulative incidence functions for competing risks data using inverse probability of treatment weighting. *Computer Methods and Programs in Biomedicine*, *129*, 63–70.

Nevitt, M. C., Felson, D. T., & Lester, G. (2006). *The osteoarthritis initiative. Protocol for the cohort study*. OAI Protocol. https://nda.nih.gov/static/docs/StudyDesignProtocolAndAppendices.pdf

Nießl, A., Allignol, A., Beyersmann, J., & Mueller, C. (2023). Statistical inference for state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring. *Econometrics and Statistics*, *25*, 110–124.

Ozenne, B. M. H., Scheike, T. H., Staerk, L., & Gerds, T. A. (2020). On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biometrical Journal*, *62*(3), 751–763.

Ozenne, B. M. H., Sørensen, A. L., Scheike, T. H., Torp-Pedersen, C., & Gerds, T. A. (2017). riskRegression: Predicting the risk of an event using cox regression models. *The R Journal*, *9*(2), 440–460.

Prentice, R. L., & Kalbfleisch, J. D. (2003). Mixed discrete and continuous Cox regression model. *Lifetime Data Analysis*, *9*(2), 195–210.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (2006). Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Statistical Science*, *21*(3), 299–309.

Rühl, J., Beyersmann, J., & Friedrich, S. (2023). General independent censoring in event-driven trials with staggered entry. *Biometrics*, *79*(3), 1737–1748.

Rühl, J., & Friedrich, S. (2024). Resampling-based confidence intervals and bands for the average treatment effect in observational studies with competing risks. *Statistics and Computing*, *34*(101), 1–13. https://doi.org/10.1007/s11222-024-10420-w

Ryalen, P. C., Stensrud, M. J., Fosså, S., & Røysland, K. (2020). Causal inference in continuous time: An example on prostate cancer therapy. *Biostatistics*, *21*(1), 172–185.

Scheike, T. H., & Zhang, M.-J. (2002). An additive-multiplicative cox-Aalen regression model. *Scandinavian Journal of Statistics*, *29*(1), 75–88.

Scheike, T. H., & Zhang, M.-J. (2008). Flexible competing risks regression modeling and goodness-of-fit. *Lifetime Data Analysis*, *14*(4), 464–483.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, *9*(6), 1187–1195.

Stensrud, M. J., Young, J. G., Didelez, V., Robins, J. M., & Hernán, M. A. (2022). Separable effects for causal inference in the presence of competing events. *Journal of the American Statistical Association*, *117*(537), 175–183.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistitcs*, *9*(1), 93–108.

U.S. Department of Health and Human Services (DHHS). (2008). 2008 Physical activity guidelines for Americans. https://health.gov/our-work/nutrition-physical-activity/physical-activity-guidelines/previous-guidelines/2008-physical-activity-guidelines

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Vansteelandt, S., Dukes, O., Van Lancker, K., & Martinussen, T. (2022). Assumption-lean cox regression. *Journal of the American Statistical Association*, *119*(545), 475–484.

Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., & Hernán, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*, *39*, 1199–1236.

How to cite this article: Rühl, J., & Friedrich, S. (2024). Asymptotic properties of resampling-based processes for the average treatment effect in observational studies with competing risks. *Scandinavian Journal of Statistics*, *51*(4), 1506–1532. https://doi.org/10.1111/sjos.12714

APPENDIX

A.1 PROOFS OF LEMMAS 2, 3, AND 4

Proof of Lemma 2. Because N_{ki} jumps at most once,

$$\max_{1\le i\le n}\left|\int_0^{t_r}\hat{\tilde{H}}_{k1}(u,t_r)\,\frac{\mathrm{d}N_{ki}(u)}{\sqrt{n}\,S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\right| < \left(\exp\left(\hat{\beta}_{kA}\right)+1\right)\max_{1\le i\le n}\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{i}}\right)\frac{1}{\sqrt{n}\inf\limits_{u\in[0,t_r]}S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}.$$

Recall that on $\mathcal{B}\times[0,\tau]$, $S^{(0)}$ converges uniformly to $s^{(0)}$, which is bounded away from zero, and that $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ is strongly consistent. For that reason, the expression above converges to 0 $\forall t_r\in\{t_1,\dots,t_l\}$ almost surely as $n\to\infty$.

In addition,

$$\max_{1\le i\le n}\left|\int_0^{\tau}\frac{1}{\sqrt{n}}\left(\hat{\tilde{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_r)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\mathrm{d}N_{ki}(u)\right|$$

$$\le\frac{1}{\sqrt{n}}\max_{1\le i\le n}\left\{\frac{1}{n}\sum_{j_1=1}^{n}\frac{1}{n}\sum_{j_2=1}^{n}\int_0^{t_r}\left(\exp\left(\hat{\beta}_{kA}\right)\right.\right.$$

$$
\left|\left(\left(1,\boldsymbol{Z}_{\boldsymbol{j_1}}^T\right)-\left(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},v)\right)^T\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\int_0^{\tau}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\,\mathrm{d}N_{ki}(u)\right|
$$

$$
+\left|\left(\left(0,\boldsymbol{Z}_{\boldsymbol{j_1}}^T\right)-\left(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},v)\right)^T\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\int_0^{\tau}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\,\mathrm{d}N_{ki}(u)\right|\Bigg)
$$

$$
\left.\cdot\frac{\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{j_1}}\right)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},v)}\,\mathrm{d}N_{kj_2}(v)\right\}
$$

$$
<\frac{\max\left\{\exp\left(\hat{\beta}_{kA}\right),1\right\}\cdot\max\limits_{1\le j_1\le n}\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{j_1}}\right)}{\sqrt{n}\inf\limits_{v\in[0,t_r]}S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},v)}
$$

$$
\cdot\left(\max_{1\le i,j_1\le n}\sup_{u\in[0,\tau],\ v\in[0,t_r]}\left|\left(\left(1,\boldsymbol{Z}_{\boldsymbol{j_1}}^T\right)-\left(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},v)\right)^T\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\right|\right.
$$

$$
\left.+\max_{1\le i,j_1\le n}\sup_{u\in[0,\tau],\ v\in[0,t_r]}\left|\left(\left(0,\boldsymbol{Z}_{\boldsymbol{j_1}}^T\right)-\left(\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},v)\right)^T\right)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\right|\right).
$$

Using the previous considerations and the fact that $s^{(1)}$ and $s^{(2)}$ are bounded on $\mathcal{B}\times[0,\tau]$ (Fleming & Harrington, 2005, Sec. 8.4), we can also conclude that the above maximum vanishes $\forall t_r\in\{t_1,\dots,t_l\}$ as n tends to ∞, which implies condition (i).

Moreover, for time points t_r and t_s with $0\le t_r\le t_s\le\tau$,

$$
\begin{aligned}
\sum_{i=1}^n X_{n,i}^{(k)}(t_r)\,X_{n,i}^{(k)}(t_s)=\;&\frac{1}{n}\sum_{i=1}^n\int_0^{t_r}\hat{\hat{H}}_{k1}(u,t_r)\,\hat{\hat{H}}_{k1}(u,t_s)\,\frac{\mathrm{d}N_{ki}(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^{t_r}\hat{\hat{H}}_{k1}(u,t_r)\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_s)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^{t_s}\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_r)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\hat{\hat{H}}_{k1}(u,t_s)\,\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^{\tau}\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_r)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\\
&\cdot\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_s)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\,\mathrm{d}N_{ki}(u),
\end{aligned}\tag{A1}
$$

as N_{ki} is a one-jump process. The first term of Equation (A1) equals

$$
\frac{1}{n}\sum_{i=1}^n\int_0^{t_r}\hat{\hat{H}}_{k1}(u,t_r)\,\hat{\hat{H}}_{k1}(u,t_s)\,\frac{\mathrm{d}M_{ki}(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}+\int_0^{t_r}\hat{\hat{H}}_{k1}(u,t_r)\,\hat{\hat{H}}_{k1}(u,t_s)\,\frac{\mathrm{d}\Lambda_{0k}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}.
$$

Due to the strong consistency of $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ and $\hat{\Lambda}_{0k}$, $\hat{\hat{H}}_{k1}$ is uniformly consistent, and so is $S^{(0)}$ on $\mathcal{B}\times[0,\tau]$ (with estimand $s^{(0)}$, which is bounded away from zero). It follows by application of the martingale central limit theorem that the first summand of the expression above converges to zero as $n\to\infty$. Using the same arguments on the remaining terms in Equation (A1), we obtain

$$\sum_{i=1}^{n} X_{n,i}^{(k)}(t_r)\, X_{n,i}^{(k)}(t_s) \xrightarrow{P} \int_0^{t_r} \tilde{H}_{k1}(u,t_r)\, \tilde{H}_{k1}(u,t_s)\, \frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}$$
$$+ \int_0^{t_r} \tilde{H}_{k1}(u,t_r)\big(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_s)\big)^T \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\big(\boldsymbol{s}^{(1)}(\boldsymbol{\beta_{0k}},u) - \boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\, s^{(0)}(\boldsymbol{\beta_{0k}},u)\big)$$
$$\times \frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}$$
$$+ \int_0^{t_s} \big(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_r)\big)^T \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\big(\boldsymbol{s}^{(1)}(\boldsymbol{\beta_{0k}},u) - \boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\, s^{(0)}(\boldsymbol{\beta_{0k}},u)\big)\tilde{H}_{k1}(u,t_s)$$
$$\times \frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)}$$
$$+ \big(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_r)\big)^T \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\int_0^{\tau}\left(\boldsymbol{s}^{(2)}(\boldsymbol{\beta_{0k}},u) - \boldsymbol{s}^{(1)}(\boldsymbol{\beta_{0k}},u)\big(\boldsymbol{e}(\boldsymbol{\beta_{0k}},u)\big)^T\right) \mathrm{d}\Lambda_{0k}(u)\right)$$
$$\times \big(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\big)^T \tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_s)$$
$$= \int_0^{t_r} \tilde{H}_{k1}(u,t_r)\, \tilde{H}_{k1}(u,t_s)\, \frac{\mathrm{d}\Lambda_{0k}(u)}{s^{(0)}(\boldsymbol{\beta_{0k}},u)} + \big(\tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_r)\big)^T \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{k}}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\, \tilde{\boldsymbol{H}}_{\boldsymbol{k2}}(t_s),$$

and thus, condition (ii) follows. □

Proof of Lemma 3.

$$\sqrt{n}\max_{1\le i\le n}\left|X_{n,i}^{(k)}(t_s) - X_{n,i}^{(k)}(t_r)\right|$$
$$\le \max_{1\le i\le n}\left\{\int_0^{t_s}\left|\hat{\tilde{H}}_{k1}(u,t_s) - \mathbb{1}\{u\le t_r\}\cdot \hat{\tilde{H}}_{k1}(u,t_r)\right| \frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\right.$$
$$\left. + \int_0^{\tau}\left|\left(\hat{\tilde{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_s) - \hat{\tilde{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_r)\right)^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\right| \mathrm{d}N_{ki}(u)\right\}$$
$$< \frac{2\big(\exp\big(\hat{\beta}_{kA}\big)+1\big)\, \max_{1\le i\le n}\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{i}}\right)}{\inf_{u\in[0,\tau]} S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}$$
$$+ \max_{1\le i\le n}\ \sup_{u,t_r,t_s\in[0,\tau]}\left|\left(\hat{\tilde{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_s) - \hat{\tilde{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_r)\right)^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T - \boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\right|,$$

i.e. $\sqrt{n}\ \max_{1\le i\le n}\left|X_{n,i}^{(k)}(t_s) - X_{n,i}^{(k)}(t_r)\right| \in O_p(1)$ (cf. the proof of Lemma 2).

Proof of Lemma 4. Using condition (iv) of Theorem 3, one can show that the expectation in Lemma 4 has the upper bound

$$\max_{1\le j\le n}\mathbb{E}\left(G_j^4\middle|\mathscr{F}_\tau\right)\sum_{i=1}^{n}\left(X_{n,i}^{(k)}(t_r) - X_{n,i}^{(k)}(t_q)\right)^2\left(X_{n,i}^{(k)}(t_s) - X_{n,i}^{(k)}(t_r)\right)^2$$
$$+ 2\max_{1\le j_1\le n}\left|\mathbb{E}\left(G_{j_1}^3\middle|\mathscr{F}_\tau\right)\right|\max_{1\le j_2\le n}\left|\mathbb{E}\left(G_{j_2}\middle|\mathscr{F}_\tau\right)\right|\sum_{i_1=1}^{n}\left(X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right)^2\left|X_{n,i_1}^{(k)}(t_s) - X_{n,i_1}^{(k)}(t_r)\right|$$
$$\cdot\sum_{i_2=1}^{n}\left|X_{n,i_2}^{(k)}(t_s) - X_{n,i_2}^{(k)}(t_r)\right|$$

$$+ 2 \max_{1\le j_1\le n}\left|\mathbb{E}\left(G_{j_1}\middle|\mathscr{F}_\tau\right)\right| \max_{1\le j_2\le n}\left|\mathbb{E}\left(G_{j_2}^3\middle|\mathscr{F}_\tau\right)\right| \sum_{i_1=1}^{n}\left|X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right|$$

$$\cdot \sum_{i_2=1}^{n}\left|X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right|\left(X_{n,i_2}^{(k)}(t_s) - X_{n,i_2}^{(k)}(t_r)\right)^2$$

$$+ \max_{1\le j\le n}\left(\mathbb{E}\left(G_j^2\middle|\mathscr{F}_\tau\right)\right)^2 \sum_{i_1=1}^{n}\left(X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right)^2 \sum_{i_2=1}^{n}\left(X_{n,i_2}^{(k)}(t_s) - X_{n,i_2}^{(k)}(t_r)\right)^2$$

$$+ 2\max_{1\le j\le n}\left(\mathbb{E}\left(G_j^2\middle|\mathscr{F}_\tau\right)\right)^2 \left(\sum_{i=1}^{n}\left|X_{n,i}^{(k)}(t_r) - X_{n,i}^{(k)}(t_q)\right|\left|X_{n,i}^{(k)}(t_s) - X_{n,i}^{(k)}(t_r)\right|\right)^2$$

$$+ \max_{1\le j_1\le n}\mathbb{E}\left(G_{j_1}^2\middle|\mathscr{F}_\tau\right) \max_{1\le j\le n}\left(\mathbb{E}\left(G_j\middle|\mathscr{F}_\tau\right)\right)^2 \sum_{i_1=1}^{n}\left(X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right)^2$$

$$\times \left(\sum_{i=1}^{n}\left|X_{n,i}^{(k)}(t_s) - X_{n,i}^{(k)}(t_r)\right|\right)^2$$

$$+ 4\max_{1\le j_2\le n}\mathbb{E}\left(G_{j_2}^2\middle|\mathscr{F}_\tau\right) \max_{1\le j\le n}\left(\mathbb{E}\left(G_j^2\middle|\mathscr{F}_\tau\right)\right)^2 \sum_{i_1=1}^{n}\left|X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right|$$

$$\times \sum_{i_2=1}^{n}\left|X_{n,i_2}^{(k)}(t_r) - X_{n,i_2}^{(k)}(t_q)\right|\left|X_{n,i_2}^{(k)}(t_s) - X_{n,i_2}^{(k)}(t_r)\right| \cdot \sum_{i_3=1}^{n}\left|X_{n,i_3}^{(k)}(t_s) - X_{n,i_3}^{(k)}(t_r)\right|$$

$$+ \max_{1\le j\le n}\left(\mathbb{E}\left(G_j\middle|\mathscr{F}_\tau\right)\right)^2 \max_{1\le j_3\le n}\mathbb{E}\left(G_{j_3}^2\middle|\mathscr{F}_\tau\right)\left(\sum_{i=1}^{n}\left|X_{n,i}^{(k)}(t_r) - X_{n,i}^{(k)}(t_q)\right|\right)^2 \sum_{i_3=1}^{n}\left(X_{n,i_3}^{(k)}(t_s) - X_{n,i_3}^{(k)}(t_r)\right)^2$$

$$+ \max_{1\le j\le n}\left(\mathbb{E}\left(G_j\middle|\mathscr{F}_\tau\right)\right)^4 \left(\sum_{i_1=1}^{n}\left|X_{n,i_1}^{(k)}(t_r) - X_{n,i_1}^{(k)}(t_q)\right|\right)^2 \left(\sum_{i_2=1}^{n}\left|X_{n,i_2}^{(k)}(t_s) - X_{n,i_2}^{(k)}(t_r)\right|\right)^2,$$

which we denote by ($A2$). According to the proof of Lemma 2, the first term can (informally) be expressed as

$$\max_{1\le j\le n}\mathbb{E}\left(G_j^4\middle|\mathscr{F}_\tau\right)\ \frac{1}{n^2}\sum_{i=1}^{n}\left(\int_0^{t_r} \mathrm{d}N_{ki}(u)\cdot O_p(1) + \int_0^{\tau} \mathrm{d}N_{ki}(u)\cdot O_p(1)\right)^2$$
$$\times \left(\int_0^{t_s} \mathrm{d}N_{ki}(u)\cdot O_p(1) + \int_0^{\tau} \mathrm{d}N_k i(u)\cdot O_p(1)\right)^2,$$

which may be further reduced to $\max_{1\le j\le n}\mathbb{E}\left(G_j^4\middle|\mathscr{F}_\tau\right)\ \frac{1}{n}\cdot O_p(1)$, as N_{ki} is a one-jump process. The term at hand is therefore negligible if $n\to\infty$ due to condition (iii) of the theorem.

Furthermore, the second and third summands in (A2) have the upper bound

$$\max_{1\le j_1\le n}\left|\mathbb{E}\left(G_{j_1}^3\middle|\mathscr{F}_\tau\right)\right| \max_{1\le j_2\le n}\left|\mathbb{E}\left(G_{j_2}\middle|\mathscr{F}_\tau\right)\right| \max_{(t_o,t_p)\in\{(t_q,t_r),(t_r,t_s)\}}\left(\sum_{i=1}^{n}\left(X_{n,i}^{(k)}(t_p) - X_{n,i}^{(k)}(t_o)\right)^2\right)^{3/2}$$
$$\times \sqrt{n}\ O_p(n^{-1/2}),$$

as a consequence of the Cauchy–Schwarz inequality and Lemma 3. With condition (i) as well as a combination of Jensen's inequality and condition (iii), we eventually obtain the representation

$$\max_{(t_o,t_p)\in\{(t_q,t_r),(t_r,t_s)\}}\left(\sum_{i=1}^{n}\left(X_{n,i}^{(k)}(t_p)-X_{n,i}^{(k)}(t_o)\right)^2\right)^{3/2} O_p(1).$$

This expression turns out to be a general upper bound for the expectation in Lemma 4 by application of similar considerations, involving the Cauchy–Schwarz inequality, Lemma 3 and the conditions of Theorem 3, to the remaining terms in (A2). Note that the $O_p(1)$ term does not depend on t_q, t_r, t_s!

For $(t_o, t_p) \in \{(t_q, t_r), (t_r, t_s)\}$, it thus remains to show that

$$\sum_{i=1}^{n}\left(X_{n,i}^{(k)}(t_p)-X_{n,i}^{(k)}(t_o)\right)^2 \le \left(L_n^{(k)}(t_s)-L_n^{(k)}(t_q)\right) O_p(1).$$

The inequality $(a+b)^2 \le 2a^2+2b^2$, $a, b \in \mathbb{R}$, suggests that

$$\begin{aligned}\sum_{i=1}^{n}\left(X_{n,i}^{(k)}(t_p)-X_{n,i}^{(k)}(t_o)\right)^2 \le \frac{2}{n}\sum_{i=1}^{n}\Bigg(&\left(\int_0^{t_p}\hat{\hat{H}}_{k1}(u,t_p)\,\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}-\int_0^{t_o}\hat{\hat{H}}_{k1}(u,t_0)\,\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\right)^2\\ &+\left(\int_0^{\tau}\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_p)-\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_o)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\\ &\times\left(\left(A_i,\boldsymbol{Z}_{\boldsymbol{i}}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\mathrm{d}N_{ki}(u)\right)^2\Bigg).\end{aligned} \tag{A3}$$

Due to the definition of $\hat{\hat{H}}_{k1}$, the first summand in (A3) has the upper bound

$$\begin{aligned}&\frac{2}{n}\sum_{i=1}^{n}\Bigg(\frac{1}{n}\sum_{j=1}^{n}\left(2\exp\left(\hat{\beta}_{kA}\right)+2\right)\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{j}}\right)\int_{t_o}^{t_p}\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\\ &\quad+\frac{1}{n}\sum_{j=1}^{n}\left(\exp\left(\hat{\beta}_{kA}\right)+1\right)\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{j}}\right)\int_{t_o}^{t_p}\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\\ &\quad+\frac{1}{n}\sum_{j=1}^{n}\Bigg(\left(\hat{F}_1\left(t_p\middle|A=1,\boldsymbol{Z}_{\boldsymbol{j}}\right)-\hat{F}_1\left(t_o\middle|A=1,\boldsymbol{Z}_{\boldsymbol{j}}\right)\right)\exp\left(\hat{\beta}_{kA}\right)\\ &\quad+\left(\hat{F}_1\left(t_p\middle|A=0,\boldsymbol{Z}_{\boldsymbol{j}}\right)-\hat{F}_1\left(t_o\middle|A=0,\boldsymbol{Z}_{\boldsymbol{j}}\right)\right)\Bigg)\\ &\quad\cdot\exp\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{j}}\right)\int_0^{t_o}\frac{\mathrm{d}N_{ki}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}\Bigg)^2\\ &\le\frac{2}{n}\sum_{i=1}^{n}\Bigg(\frac{2}{n}\sum_{j=1}^{n}\left(9\exp\left(2\hat{\beta}_{kA}\right)+18\exp\left(\hat{\beta}_{kA}\right)+9\right)\exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z}_{\boldsymbol{j}}\right)\int_{t_q}^{t_s}\frac{\mathrm{d}N_{ki}(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}\end{aligned}$$

$$+\frac{2}{n}\sum_{j=1}^{n}\Bigg(2\left(\left(\hat{F}_1\left(t_s\middle|A=1,\boldsymbol{Z_j}\right)\right)^2-\left(\hat{F}_1\left(t_q\middle|A=1,\boldsymbol{Z_j}\right)\right)^2\right)\exp\left(2\hat{\beta}_{kA}\right)$$

$$+2\left(\left(\hat{F}_1\left(t_s\middle|A=0,\boldsymbol{Z_j}\right)\right)^2-\left(\hat{F}_1\left(t_q\middle|A=0,\boldsymbol{Z_j}\right)\right)^2\right)\exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_j}\right)\int_0^{\tau}\frac{\mathrm{d}N_{ki}(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}\Bigg). \tag{A4}$$

For the last step, we used again that $(a+b)^2 \le 2a^2+2b^2$, as well as the Cauchy–Schwarz inequality and $(a-b)^2 \le a^2-b^2$ for $0 \le b \le a$.

The second summand in (A3) is further equal to

$$\frac{2}{n}\sum_{i=1}^{n}\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_p)-\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_o)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_p)-\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_o)\right),$$

because $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}$ is symmetric and N_{ki} is a one-jump process. Note that

$$\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_p)-\hat{\hat{\boldsymbol{H}}}_{\boldsymbol{k2}}(t_o)=\frac{1}{n}\int_{t_o}^{t_p}\left(\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{A=1}}(u,t_p)-\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{A=0}}(u,t_p)\right)\sum_{i=1}^{n}\frac{\mathrm{d}N_{k,i}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)}$$
$$-\frac{1}{n}\int_0^{t_o}\left(\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{A=1}}(u,t_o,t_p)-\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{A=0}}(u,t_o,t_p)\right)\sum_{i=1}^{n}\frac{\mathrm{d}N_{k,i}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)},$$

with

$$\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{a}}(\boldsymbol{u},\boldsymbol{t})=\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}\{k=1\}\cdot\hat{S}\left(u-\middle|a,\boldsymbol{Z_i}\right)-\hat{F}_1\left(t\middle|a,\boldsymbol{Z_i}\right)+\hat{F}_1\left(u\middle|a,\boldsymbol{Z_i}\right)\right)$$
$$\times\left(\left(a,\boldsymbol{Z_i}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\cdot\exp\left(\hat{\beta}_{kA}\cdot a+\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right),$$

$$\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{a}}(\boldsymbol{u},\boldsymbol{s},\boldsymbol{t})=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{F}_1\left(t\middle|a,\boldsymbol{Z_i}\right)-\hat{F}_1\left(s\middle|a,\boldsymbol{Z_i}\right)\right)\left(\left(a,\boldsymbol{Z_i}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\exp\left(\hat{\beta}_{kA}\cdot a+\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_i}\right).$$

Besides, the product $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}$ is positive definite because of the definitions of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}$. Since

$$(\boldsymbol{a}-\boldsymbol{b})^T\boldsymbol{A}\,(\boldsymbol{a}-\boldsymbol{b})\le 2\,\boldsymbol{a}^T\boldsymbol{A}\,\boldsymbol{a}+2\,\boldsymbol{b}^T\boldsymbol{A}\,\boldsymbol{b}\qquad\text{and}\qquad\left(\sum_{i_1=1}^{n}\boldsymbol{a}_{\boldsymbol{i_1}}\right)^T\boldsymbol{A}\left(\sum_{i_2=1}^{n}\boldsymbol{a}_{\boldsymbol{i_2}}\right)\le n\sum_{i=1}^{n}\boldsymbol{a}_{\boldsymbol{i}}^T\boldsymbol{A}\,\boldsymbol{a}_{\boldsymbol{i}},$$

for a positive (semi-)definite matrix $\boldsymbol{A}$ and vectors $\boldsymbol{a},\boldsymbol{b},\boldsymbol{a_1},\dots,\boldsymbol{a_n}$, the second summand in (A3) has the upper bound

$$\frac{2}{n}\sum_{i=1}^{n}\Bigg(\frac{2}{n}\int_{t_o}^{t_p}\left(\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{A=1}}(u,t_p)-\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{A=0}}(u,t_p)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{A=1}}(u,t_p)-\hat{\chi}_{\boldsymbol{k1},\ \boldsymbol{A=0}}(u,t_p)\right)$$
$$\times\frac{\mathrm{d}N_k(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}+\frac{2}{n}\int_0^{t_o}\left(\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{A=1}}(u,t_o,t_p)-\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{A=0}}(u,t_o,t_p)\right)^T$$
$$\times\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{A=1}}(u,t_o,t_p)-\hat{\chi}_{\boldsymbol{k2},\ \boldsymbol{A=0}}(u,t_o,t_p)\right)\cdot\frac{\mathrm{d}N_k(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}\Bigg)$$

$$\leq \frac{2}{n}\sum_{i=1}^{n}\left(\frac{4}{n^2}\sum_{j=1}^{n}\exp\left(2\hat{\beta}_{kA}+2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_j}\right)\int_{t_q}^{t_s}\left(\left(1,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T\right.$$

$$\times\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(1,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\frac{\mathrm{d}N_k(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}$$

$$+\frac{4}{n^2}\sum_{j=1}^{n}\exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_j}\right)\int_{t_q}^{t_s}\left(\left(0,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T$$

$$\times\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(0,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\frac{\mathrm{d}N_k(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}$$

$$+\frac{4}{n^2}\sum_{j=1}^{n}\left(\left(\hat{F}_1\left(t_s\middle|A=1,\boldsymbol{Z_j}\right)\right)^2-\left(\hat{F}_1\left(t_q\middle|A=1,\boldsymbol{Z_j}\right)\right)^2\right)\exp\left(2\hat{\beta}_{kA}+2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_j}\right)$$

$$\cdot\int_0^{\tau}\left(\left(1,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(1,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\frac{\mathrm{d}N_k(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}$$

$$+\frac{4}{n^2}\sum_{j=1}^{n}\left(\left(\hat{F}_1\left(t_s\middle|A=0,\boldsymbol{Z_j}\right)\right)^2-\left(\hat{F}_1\left(t_q\middle|A=0,\boldsymbol{Z_j}\right)\right)^2\right)\exp\left(2\hat{\boldsymbol{\beta}}_{\boldsymbol{kZ}}^T\boldsymbol{Z_j}\right)$$

$$\left.\cdot\int_0^{\tau}\left(\left(0,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\left(\left(0,\boldsymbol{Z_j}^T\right)^T-\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)\frac{\mathrm{d}N_k(u)}{\left(S^{(0)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}},u)\right)^2}\right).$$

For the last two terms, we used that for $0 \leq b \leq a$, $(a-b)^2 \leq a^2 - b^2$. Note also that it is possible to extend the integral limits here because of the positive definiteness of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{ki}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}^{-1}$.

Together with (A4), it follows finally that

$$\sum_{i=1}^{n}\left(X_{n,i}^{(k)}(t_p)-X_{n,i}^{(k)}(t_o)\right)^2 \leq 36\left(L_n^{(k)}(t_s)-L_n^{(k)}(t_1)\right),$$

$\forall k \in \{1,\dots,K\}$. □

TABLE A1 Summary of the OAI data subset analyzed.

Covariate	DHHS aerobic activity guideline attained: No ($n = 414$)	DHHS aerobic activity guideline attained: Yes ($n = 47$)
Site		
A	58 (14.01%)	9 (19.15%)
B	104 (25.12%)	10 (21.28%)
C	147 (35.51%)	17 (36.17%)
D	69 (16.67%)	10 (21.28%)
E	36 (8.70%)	1 (2.13%)
Sex: male	191 (46.14%)	34 (72.34%)
Age, mean (SD)	68.08 (8.56)	63.74 (7.67)
BMI, mean (SD)	29.59 (4.80)	26.43 (2.85)
Pack-years of smoking, mean (SD)	10.29 (18.07)	10.43 (17.00)
Race: White	335 (80.92%)	44 (93.62%)
Education: college degree	254 (61.35%)	40 (85.11%)
Comorbidity score		
0	270 (65.22%)	35 (74.47%)
1	86 (20.77%)	4 (8.51%)
2	31 (7.49%)	5 (10.64%)
3–6	27 (6.52%)	3 (6.38%)
Average knee pain over the last week		
0	116 (28.02%)	23 (48.94%)
1–3	161 (38.89%)	14 (29.79%)
4–6	105 (25.36%)	8 (17.02%)
7–10	32 (7.73%)	2 (4.26%)
Kellgren–Lawrence grade: 4	135 (32.61%)	12 (25.53%)
Prior knee surgery	166 (40.10%)	24 (51.06%)
Average accelerometer count per day	180,829 (82,884)	370,055 (113,784)

Abbreviations: DHHS, US Department of Health and Human Services; OAI, Osteoarthritis Initiative.